VOLUME 53
No. 3

WHOLE NO. 240
1941

Psychological Monographs

EDITED BY
JOHN F. DASHIELL
UNIVERSITY OF NORTH CAROLINA

The Effects of Praise and Blame as Incentives to Learning

By

HERMANN O. SCHMIDT

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PUBLICATIONS OFFICE
NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

ACKNOWLEDGMENTS

THE writer acknowledges with deep appreciation his debt of gratitude to Professor Florence E. Bamberger for much helpful criticism and objective assistance throughout the course of his training. He is grateful to the chairman of his committee, Dr. J. M. Stephens, for his whole-hearted guidance, constructive and penetrating criticism, and helpful advice during associations with him and through the course of this present study. He wishes to acknowledge his debt to his friend, Dr. G. Wilson Shaffer, without whose counseling and timely material assistance this study would never have become *un fait accompli.* He also wishes to express his thankfulness to Dr. Wendell W. Cruze and to Dr. Arthur L. Lichtenstein, of his committee, for valuable criticism and suggestion.

He is indebted to Dr. John W. Fertig for suggestions and assistance with the statistical material, and to Dr. Sydney M. Newhall for many pleasant and fruitful hours spent together. He is also appreciative of the assistance given by Captain Frank H. Forney, C.E., in photographing the apparatus, and by Captain Roland D. Foley, F.A., for reproducing the many charts and tables; and to his sister, Virginia, for much patience and understanding.

He wishes to express his appreciation to Major Louis E. Lamborn, headmaster, and Dr. T. H. Wilson, head of the Upper School, McDonogh School, McDonogh, Maryland; and to Mr. Wilson P. Sperow, principal of the Woodland Way Junior High School, Hagerstown, Maryland, for their interested and willing coöperation in permitting the tests to be given at their schools.

It is a pleasure to acknowledge special indebtedness to Miss Clara P. McMahon, "Tester B" of this study, for her most generous assistance throughout this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

THE EFFECTS OF PRAISE AND BLAME AS INCENTIVES TO LEARNING

CHAPTER I

The Problem

Introduction

THE purpose of this investigation has been to study the effectiveness of praise and blame as incentives to learning; and further, to examine into the possible influence of different testers and emotional factors that may have been playing a part. The method adopted was to run parallel experiments with two testers, though for practical reasons there was no duplication of the emotional investigation. There were two experimental groups in each case, a Praise and a Blame group, and a Control group. The answer to the question of the effectiveness of praise and blame was sought through a comparison of the gains made by each group from an initial test, when no incentives were administered, to a final test, where all groups received a common incentive.

Virtually all of the previous experimental work that has been done with humans on incentives in learning, aside from being so varied as to leave nothing but confusion in its wake, has missed the vital point that in measuring the amount of learning taking place, it has been measuring *output* of some sort, rather than *acquisition*.

This problem has been recognized in the infra-human studies by Finan (30), Leeper (55, 56), Lashley (54), Tolman (74), Tolman, Hall & Bretnall (75), Blodgett (15), and others.

For example, Blodgett, working with control and experimental groups of rats running mazes, sought to study the efficiency of units of practice when unaccompanied by reward. The experimental group received no reward during the first part of learning, while the control group was rewarded throughout the whole of learning. His findings were, in brief:

1. Rats run under a non-reward condition learned more slowly than rats run under a reward condition. This held both for errors and time.
2. Rats previously run under non-reward conditions made a great improvement when suddenly rewarded. On the first day after the reward their drop in mean error was greater than that made by control on any single day.

3. This would seem to indicate that non-rewarded rats were developing a *latent learning* which they could make use of when the reward was introduced.

No such clear-cut picture exists in the human field, as a review of the literature shows.

Historical

Any real review of the history of motivation would demand far more space than can logically be assigned here, for the history of motivation is a history of emotions and instincts. This has been pointed out by Brenner (18), who presented a brief résumé running from Aristotle to the date of his study, 1934, with detailed attention being given to the effects of comments as incentives. His bibliography lists twenty titles. He brings out the natural transition from the experiments with animals, with Thorndike the leading investigator, to experiments in the human field.

Brenner's review of the studies employing comments as incentives begins with the work of Kirby (48) in 1913, then discusses in order, Gilchrist (36), Chapman and Feder (21), Gates and Rissland (35), Hurlock (41, 42), Sims (69), Symonds and Chase (72), Deputy (26), and Maller (59). He finds generally that the shortcomings of these studies (in which we concur) may be listed as follows:

1. Small number of cases.
2. Inadequacy of the grouping of the subjects to make them comparable.
3. No attempt to define and isolate an incentive as a single, specific variable.
4. Inadequacy of the statistical treatment of the data.
5. Interpretations not substantiated by statistical evidence.
6. Unwarranted generalization.
7. Undue expectations that a certain incentive active in a certain way in a laboratory situation will act similarly in every school situation.

He points out further that the age-range of the subjects (from 7-year-olds to adults), the varying experimental conditions, and the differences in the kinds of incentives, were all bound to bring about conflicting and incomparable results. However, all seem to agree (by implication, at least) that any change in the learning situation will bring about an increment in *performance* (this word has been italicized deliberately; it is crucial to the question and will be developed more fully presently). Brenner makes a plea for a clearer formulation which will provide a more definite goal to which research can be directed.

It appears, indeed, that this latter has been one of the chief factors making for a confusion of issues and a lack of comparableness in the larger majority of the studies. It is unfortunate also that the terminology is no more clear-cut than it is. Praise vs. Blame, Reward vs. Punishment, Knowledge of Results vs. No Knowledge of Results, have in many instances been examining the same thing, or have been interwoven in the same work. These terms have been in large measure not nearly so definable as their connotations suggest; nor have they always been truly indicative of their investigation. For example, we might cite Binet and Vaschide (10), after Hurlock (45): *Expérience de forces musculaires et du fond chez les jeunes garçons.* Actually, this was an experiment in motivation with 43 school children, who were *encouraged* to break their previous scores. The encouragement, "*Allons, tu peu faire mieux que ça, toi,*" always resulted in improvement.

Regardless of the announced objective of an experiment, be it the effects of praise vs. blame, or bell-right vs. bell-wrong, the matter of knowledge of results plays a

rôle. No doubt, some of the conflicting or indecisive results have been due, in part at least, to oversight of this problem.

Judd (47), working with the projection of a straight line without knowledge of results, has shown that there is no improvement if this factor is removed. Arps (6), working with three subjects with the ergograph, found that partial knowledge of results brought about a decrement in performance. Later (7), he found that the absolute amount of work done and rate were exceeded under conditions of knowledge of results. Similarly, Elwell and Grindley (29), having subjects attempt to hit a target with a spot of light by manipulating levers with both hands, found improvement when results could be observed, no improvement or loss when target was hidden from view. However, Spencer (70), repeating Judd's experiment, with four subjects, found that there was improvement with three ot the four subjects.

Unequivocal answers to the problem of Praise vs. Blame (or similar incentives) have been beclouded also because of losing sight of the fact of other motivation possibly acting. There is a social aspect, as well as the emotional and intellectual, that may be exerting its influence. For example, it has been shown by Triplett (77), Arps (5), and others that group competition is more effective than competition with oneself. However, Maller and Zubin (60), employing a "very strong incentive of rivalry," failed to find this true. The ages of the subjects were not comparable in these studies, nor the experimental media, and Maller and Zubin's criteria were somewhat different. They did find, however, that more items were attempted, but accuracy was decreased.

It is curious to note the widespread prevalence of the idea of the effectiveness of employing praise rather than blame. Yet there is to be found no unequivocal evidence to support this stand. In everyday life sentimental behavior virtually makes mandatory the "pat on the back," and abjures the otherwise; but it is surprising when one encounters this notion in some scientific writings, or in writings from the pens of persons with scientific training. By way of example, one reads (24):

> . . . there appear to be three main sources of motivation among children.(1) They respond better to praise than to punishment; . . .
> . . . a child cannot be forced into solving an arithmetic problem by being punished for not solving it. You must realize that punishment disorganizes a child's ability to think or work. . . . Just after they have been punished they are less integrated than usual and less able to concentrate. Punishment is therefore never an appropriate form of motivation.

Is a child any the more integrated when excited with pleasure?

Countless other citations could be made; but the point is, or appears to be, that in writing a chapter, or in drawing a conclusion, some investigators have not penetrated deeply enough. Briggs (19), in a two-page article, reports a study employing the Laird technique of subjective opinions, with "considerably more than 300 graduate students" in a summer course on "The Improvement of Instruction in Secondary Schools" at Teacher's College, Columbia. Without preparation or discussion beforehand the students were asked to report on a twenty-one item questionnaire (Laird's list plus "best liked" and "least liked" teacher) as to whether they worked better, worse, or the same in high school. He finds his results in general agreement with Laird's, and from this concludes that the evidence is convincing that commendation, praise, and encouragement are superior to censure, ridicule, threats, and punishment. He reports further in the article a study in the Speyer Experimental Junior High School in which two teachers known to be severe gave a pre-test, alternated praise and reproof, and gave a final test. Eighty-seven per cent of the pupils did better under praise than under reproof. No mention was made of the number of subjects, except

to say that there were four classes per teacher, nor of age or sex. From this he finds added evidence in favor of praise over blame.

The first systematic review of incentives is found in Hurlock (45). She reviews both the infra-human and human fields. The analysis in the human field is grouped in keeping with the trend of investigations into the incentives used rather than the situation involved. These are: (a) knowledge of results, (b) praise and reproof, (c) rewards, (d) punishment, (e) the influence of an audience, (f) rivalry, (g) distraction, and (h) music. In the field of particular interest to us here, (b), she summarizes the work of Kirby (48), Binet and Vaschide (10), Scott (67), Ream (64), Briggs (19, 20), Laird (51, 52), Gilchrist (36), Gates and Rissland (35), Hurlock (41, 42, 44), Cohen (23), and Watts (80). She draws no conclusions, nor summarizes findings, but presents merely brief synopses of the studies.

Diserens and Vaughn (27) present a comprehensive review of the field of motivation, but add no further studies to those already mentioned above in the field of social motivation (praise and blame). They conclude *(p. 52)* by re-naming Hurlock (41) that praise is more effective than blame; that indifference is detrimental. They point out that social motivation requires an approach from the quantitative standpoint in order to discover the effect of a group in various attitudes on the performance of other individuals or groups also varying in attitude. They also feel it necessary for investigation to be made of groups varying greatly in numerical range. They have posited several "laws" affecting motivation. These are:

1. The number of motives available for experimentation increases with the complexity and degree of development of the organism.
2. The energy of a motive varies directly with its primitiveness (punishment, food, sex are more energetic than social motives).
3. The degree of unity of motivating forces in any situation varies inversely with the intelligence of the motivated organism.
4. The effectiveness of a given motive in any situation varies directly with the number of coöperating motives or facilitating factors, and inversely with the number of competing motives or inhibiting factors.

Davis and Ballard (25) present a review and a tabulation of various incentives used in the school situation. They divide these incentives into three classes: intellectual, emotional, and social. The intellectual incentives refer to devices such as informing the pupil of success or failure; the emotional incentives include encouragement or discouragement; and the social incentives take in the devices of the pupil working in a social situation. It is in the second classification that we are primarily interested here. The bibliography contains eight items of which the last, for our purposes at least, properly belongs under the classification of punishment and rewards. The other studies have been mentioned before. The authors conclude: that praise is better than reproof; that both may be effective, but praise is better; that too much praise will defeat its purpose (a point that has not been brought out previously); that individual differences in pupils and teachers will govern the extent to which praise or reproof may be used; that some comment is better than none at all; that praise is better for younger children, and for the dull; boys are more influenced by reproof, girls by praise; positive incentives are better for all ages, grades, and intelligence than negative incentives.

Irwin (46) reviewed the field of motivation with a list of 86 studies. He stated the results are conflicting, but there appears a great deal of evidence to show that the effect of rewards are beneficial to learning, the degree of benefit varying with the type and the amount of rewards given. The effects of punishment seem to be more doubtful.

Trow (78) reviewed the studies of Wolf (81), Ryans (66), Benton (8), Abel (1), Anderson (4), and Sears (68) in the field of motivation which pertains here.

Young (84, *pp. 388–416*) summarized the work on motivation, but cites only four studies on praise and blame. He suggests the need exists to study the effects of praise and blame in relation to self-evaluation; *i.e.,* to the level of self-esteem.

Murphy, Murphy and Newcomb (61) in their chapter on "Aggression and Competition" consider the problem of praise and reproof as a question of the relationship between the individual personality and the situation. In a series of tables they list and digest in several divisions a large number of studies concerned with incentives. These tables contain the author, date, subjects, methods, and findings. Relative to praise and reproof, and under this division, the authors review the studies of Brenner (18), Briggs (19, 20), Gates and Rissland (35), Gilchrist (36), Hurlock (41, 42), Laird (50), 51, 52), Warden and Cohen (79), and Wood (82). Under "studies of competition" they list Leuba (57), whose study involved some use of praise. In "knowledge of results" there appear Book and Norvell (17), employing some reproof, and Chapman and Feder (21). Under "multi-incentive studies" there are reviewed Anderson (3), Anderson and Smith (4), Benton (8), Chase (22), Kirby (48), and Sullivan (71). The authors call attention to the conflicting evidence of the studies they have summarized and attribute this confusion to a failure to have considered the individual in relation to the cultural background, and his own peculiar personality make-up.

Bird (11) presents a clinical approach with some numerical enhancing on the analysis of factors influencing learning in 100 cases of healthy young children with age range of four to six years. Only average MA cases were studied, the others being rejected. The MA range is stated as approximating a normal curve. MA was measured by Stanford-Binet and the Rhode Island Intelligence Examination. The type of learning employed was the kinesthetic part (finger tracing of letters, preparatory to writing on board) of the stages in learning to read by the Henry Barnard School method. The measure was the number of errors or attempts to learn a new letter.

Of the 100 children, 30 were found with habitual personality handicaps that interfered with learning; 37 showed unmistakable affective disturbances, but to a lesser degree than the others. These were such traits as shyness, tantrums, introversion, bullying, etc.

In speaking of the "30-group," the author says of the shy, lacking-in-confidence, dislike-of-scrutiny, fear-of-task individual that close on the heels of this obstacle in learning there was excessive dependence on commendation. Sometimes this was sought to satisfy the ego, or for encouragement for further effort. Of these dependent individuals, eight were found definitely delayed in their progress whenever praise was not forthcoming.

Four children desired to win distinction by unusual behavior. To these, admonition was more satisfying than indifference.

Adams (2) presents a philosophical essay on praise. He considers it a necessary and desirable weapon in the hands of the teacher early in the school-life of the child, but feels that with maturation this must be eliminated gradually.

Forlano and Axelrod (31) made a study of 131 pupils in four classes of the fifth grade in a New York public school. These were divided into three groups—control, praise, and blame, and the latter two into two groups each, introvert and extrovert. The Woodworth-Wells Number Cancellation Test was the testing vehicle. Three weeks elapsed between tests in which a thirty second practice period was allowed, followed by a two minute test period. The pupils were called to the desk, a mark of P or G placed upon their papers according to a prearranged plan, then they were

asked to return to their seats leaving their papers face down on the desk. Announcement was made that those who received P had done poorly, while those who had received G had done well. The next test was given at once. The control group followed the general procedure but without the incentives being administered. The three groups were equated on the basis of initial scores, CA being disregarded since the correlation was only —.03. All groups gained in succeeding trials (2) over the first. There were significant differences in the second trial for the IB over the control group and of IB over IP. On the third trial differences were again significant for IB over C, EB over C, and EB over EP. On this trial the standard ratios had increased for the praise group. They concluded that blame was a more effective incentive than praise, and speculated on the increase of the standard ratios for the praise group. Would this have continued to significance had more trials been employed?

Blissenbach (14) presents an unfactual discourse closely approaching complete sentimentality on the value of praise. It is worthless as far as providing any scientific data on the question.

Blankenship and Humes (13) attempted to determine the effect of praise and reproof on memory span performance and memory span reliability. Their subjects were female college undergraduates divided into three groups: control, 44; praise, 43; and reproof, 43. The average age was about twenty years. Statements of praise or reproof were made verbally. The test media were two sets of digits varying from 5 to 13 numbers, which were read to the subjects at a rate of slightly less than one per second. Each group was tested twice, a different series being used the second time, after a week's interval. They concluded that the effects of praise or reproof were negligible, as far as memory span is concerned, the differences between gains of each group being statistically not significant. The approved group showed significantly more reliable memory span performance (0.88) than the reproved (0.41), or the control (0.47). Differences in reliability coefficients between last two groups is not significant. There was a low correlation coefficient in each group (from .07 to .12), indicating no relationship between memory span gain and age in each group.

We gather from a review of the literature that:

1. The effects of praise being a greater incentive than blame, or vice versa, is still an unsettled question.
2. The later studies tend to show little difference between the effectiveness of these incentives, and if leaning in any direction, do so towards blame.
3. Both praise and blame give increased performance on the second and third trials. Additional trials will produce one of two results, dependent upon the study chosen: either a continuing increment or a decrement in performance.
4. It appears that both praise and blame show only slightly increased performances in the second and third trials over any other change that might be introduced into the learning situation.
5. Really adequate comparisons cannot be made between studies, because:
   a. The variables of IQ, CA, MA, grade, initial ability, intensity of motivation (or type), and Emotional Aspect, have not been equated for, or taken account of, in many studies, or are on entirely differing planes.
   b. Techniques, approaches to the question, test media, vary considerably between experiments.
   c. Often the statistical treatment of the data has been unreliable or inadequate.
   d. Generally, a lack of any common ground upon which to make comparisons.
6. The actual motivating factor has been accepted too casually (with the possible single exception of Brenner):
   a. Equal statements of praise or reproof are not necessarily of equal weight in effectiveness; what might act as blame for one individual (donor or recipient) might act as praise for another.

b. The method of administering the incentive will affect its potency, and has varied considerably from experiment to experiment; from a simple statement of praise or reproof to a quite lengthy statement, and from the giving of it to the individual alone to the giving of it to the group or to an individual before a group.

7. The *total* situation, including the physiological, has been given little attention.
8. In none of the studies has the factor of the influence of the experimenter's personality been taken into account.
9. And, finally, there has been no distinction made between learning and the ends employed to seek its measurement.

Methods

The preceding review of the literature leaves no definite conclusions to be drawn as to the effectiveness of praise and blame as incentives to learning. This has been brought about, it is felt, through there being no clear-cut conception of the basic principles involved, or through a non-realization or non-differentiation of what was being sought and what was actually being done. It is true that performance has been mentioned frequently in the literature cited, but this has been done casually and interchangeably with learning.

Quite irrespective of the "system" of psychology adhered to, the fact seems patent that very little is known of the actual processes involved in the matter of learning. Experimental work (this study included) which has attempted to solve the problem of learning or of motivation in learning has had to resort to a measuring of *output* on the part of the subjects. What forces, or factors, were operating at the time of acquisition and at the time of reproduction could not be discerned. The question hinges upon two factors: the amount of learning previously acquired; and the forces operating at the time of performance.

The crucial point in this experiment comes from a comparison of the improvement of each group from the initial test, S_I, to the final test, S_e (these gains are shown in Table VIII, p. 38).[1]

In the case of S_I, it is reasonably certain that the urge to exert oneself was comparable in all the groups. In S_I, no specific incentives were applied; the subjects were simply requested to do the test. In S_e, all the groups were administered the same incentive. Now, if the energy output on S_e were comparable in the groups, it could be said that the group making the highest score had reached a higher degree of attainment, and that the group improving the most from S_I to S_e had learned the most.

[1] S with subscript refers to the stimulus-test situation. Thus S_I refers to the initial test; S_1 to the first test immediately following this; S_2 to the next; then, S_3; and S_e, the final test, the one where the common incentive is given. In addition, in the College group, S_p indicates a preliminary period to the main testing; and S_F, a final after-test period. The order was thus: $(S_p) \ldots S_I \ldots S_1 \ldots S_2 \ldots S_3 \ldots S_e \ldots (S_F) \ldots$

Any such hypotheses as above could not be ventured were the results to be taken from S_1, S_2, or S_3, the intermediate learning tests. On any of these it cannot be certain whether a high-ranking group on one of these tests has actually *learned* more or has merely *expended more energy* because of the incentive.

Discussion

Let this question be aired more fully. It has been observed that virtually nothing but confusion exists in so far as a definite answer is concerned relative to the effects of praise or blame as incentives to learning. In one study the evidence points to blame as being the more effective; in another study the data all favor praise. How does this come about? The answer lies in the design of the experiments. A test is given. This is followed by another test, or perhaps several, under incentive conditions (praise, blame, or no comment). Then curves are plotted and data assembled. From these, conclusions favoring one incentive or another are drawn. In the one case, there would be found evidence indicating praise to be the most effective; in the other, just as cogent evidence in favor of blame. But such experiments have been superficial, have not penetrated deeply enough; nor have they been set up to really answer the question: "*Is praise or is blame the more effective incentive to learning?*" What, then, have prior experiments in the human field *actually* been concerned with? Output, performance, expenditure of energy, and the like. Were the control groups *learning* any the less than the other group? The data do not present any evidence adequately to answer this; for these data have been examining *performance* and NOT *learning*. What would the control groups have done had they been stirred into action, had they been motivated into *performing* with an *equal* amount of *effort* as the other groups? The answer is not to be found in the data in the human field.

In the animal field, Blodgett (15) [2] has thrown some light on the problem. His learning curves show almost a straight line flatness for the rats running the maze without the food reward; that is, there was little decrease in errors or time scores. On the contrary, the curves of the rats given the reward at the end of each run fell off sharply. After varying periods of time (3 days to 7 days), food reward was introduced to the previously non-rewarded animals. When this was done, the error and time curves fell off precipitously, the improvement (drop in mean error) exceeding in some instances the best improvement on any single day of the always-

[2] *Op. cit.*, pp. 120–121.

rewarded group. It could be nothing but a stretching of the imagination to attribute the tremendous gain of the latterly rewarded rats to a sudden accumulation of learning. Clearly, prior to the reward these animals had been learning the maze, as Blodgett points out, but had been without any pressing motivation to hurry to the exit.

Now, look at two of Blodgett's curves.[3]

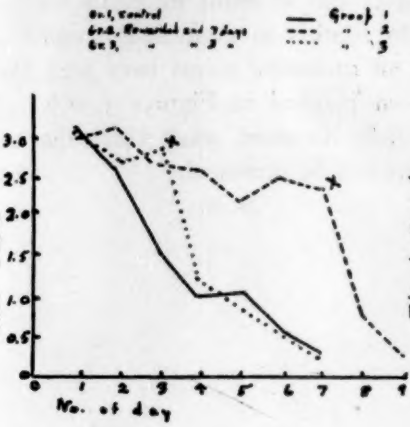

FIG. 1

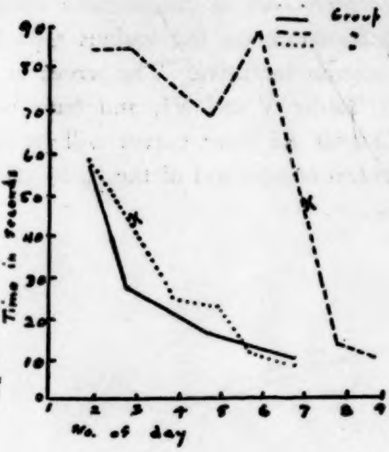

FIG. 2

Had the experiment terminated at the points *x*, there could have been but one implication to follow: rats rewarded at the end of running a maze learned more than those not rewarded. Such analysis or scrutiny would be in keeping with the majority of past experiments in human motivation. But following along with Blodgett, it is to be observed that conclusions based on such scrutiny are insufficiently founded. Thus, learning is not measured as a *thing* in itself, but as *amount of output.* Further analysis of the Blodgett curves reveals a quite different solution where *all groups have the same incentive to produce.* After the introduction of a common incentive, the previously unrewarded rats not only reduced their error and time scores pronouncedly, but in some instances improved more than the always-rewarded rats.

Logically, then, if the measure of learning must be in performance, then there must be comparable opportunities to perform for both control and experimental groups. Remove the equal opportunity for groups *to want to produce,* there must follow an entirely different measure of learning if, indeed, it is a genuine measure. At any rate, learning as evidenced through

[3] *Ibid.*, Figs. 5 and 6.

performance is only partially and incompletely measured. Even a cursory glance at the curves of the present experiment (Figures 3 to 13, pp. 11–21) will show the great number of possible conclusions that might be drawn were these to be based on the results only of the first, second, or third tests after the initial test.

However, the gains from test to test under the influence of the different incentives are of considerable interest; and it seems important to show performance on the various tests in relation to performance under the common incentive. The scores at all successive points have been given in Tables V and VI, and have been graphed in Figures 3 to 13. In Chapter III these curves will be briefly discussed, after which the more crucial comparison of the S_e–S_I gains will be presented.

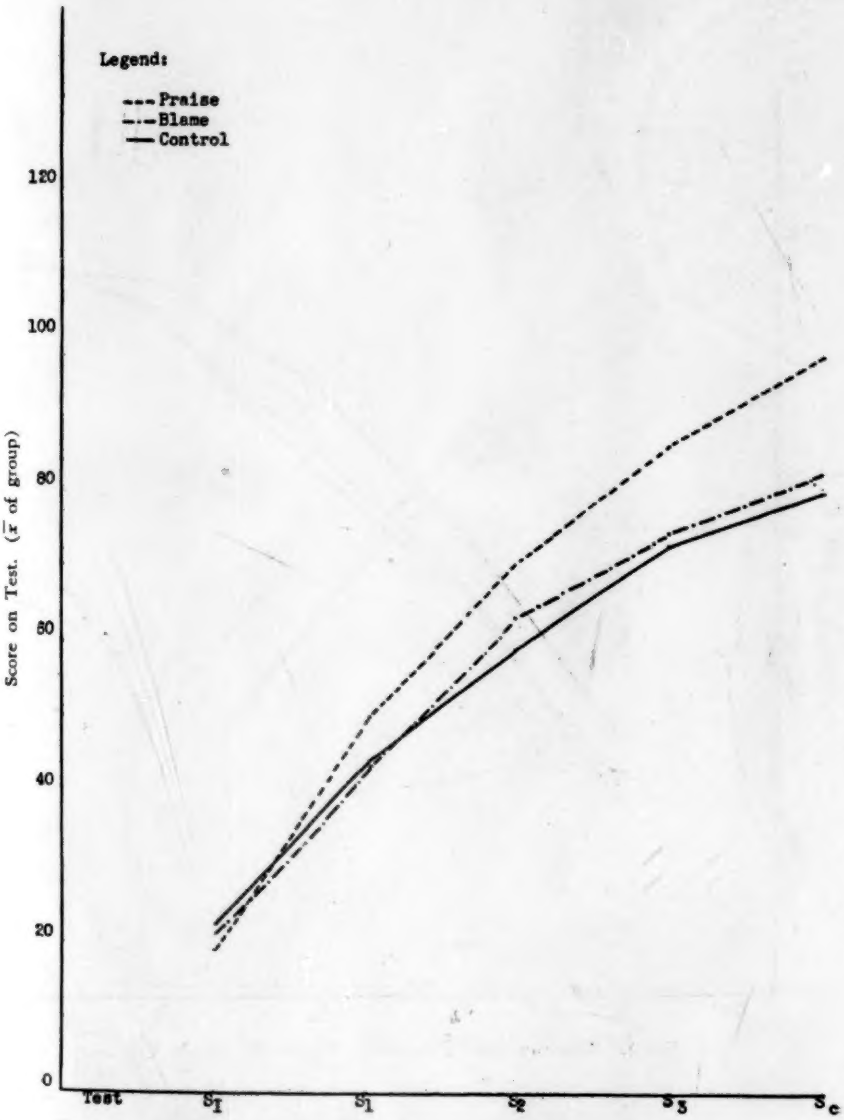

FIG. 3. BOYS AND GIRLS COMBINED. TESTER A. GRADE VII

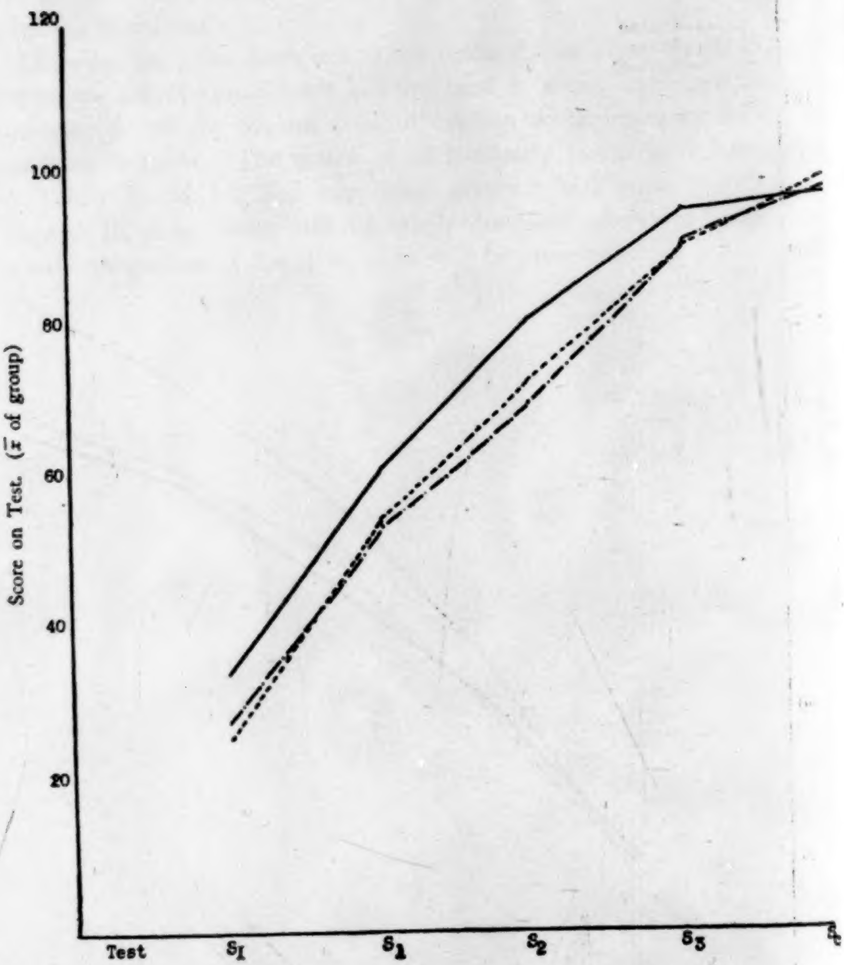

FIG. 4. BOYS AND GIRLS COMBINED. TESTER A. GRADE VIII

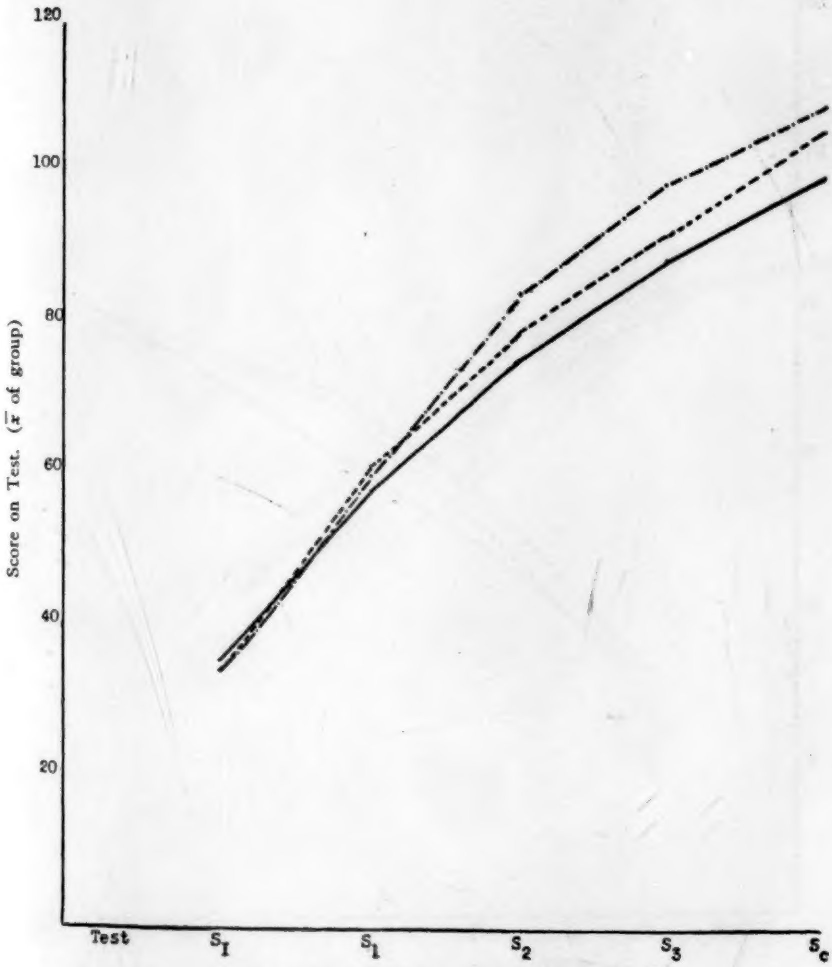

FIG. 5. ALL BOYS. TESTER A. HIGH SCHOOL

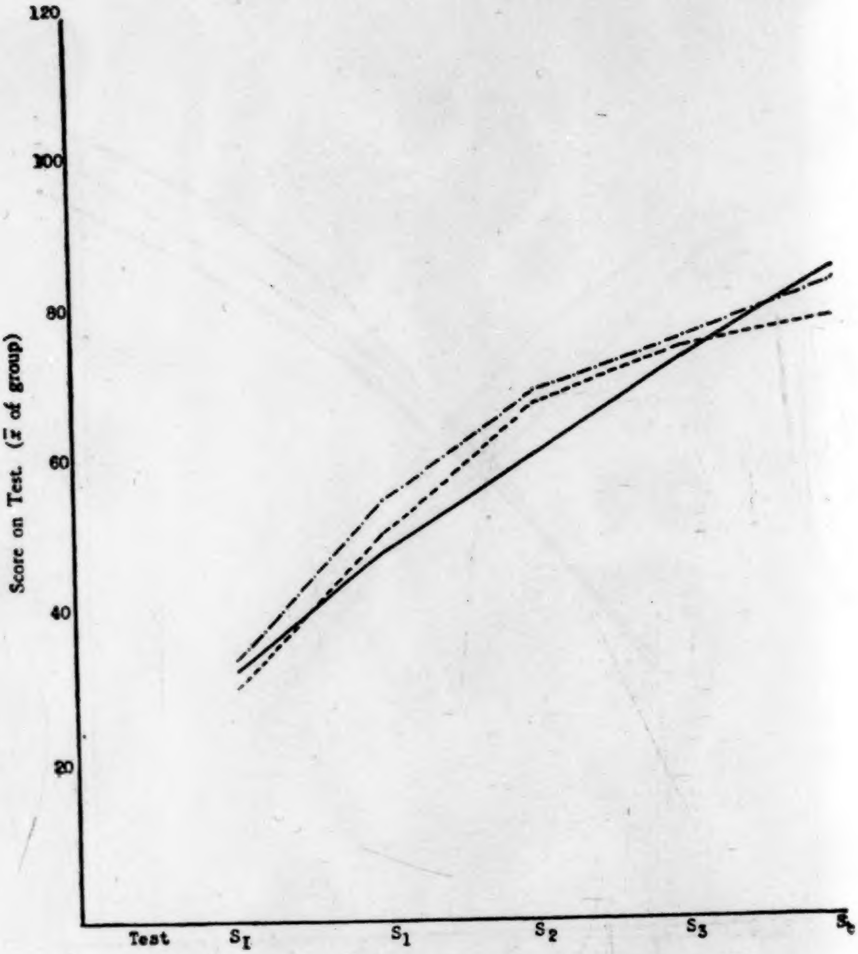

FIG. 6. BOYS AND GIRLS COMBINED. TESTER B. GRADE VII

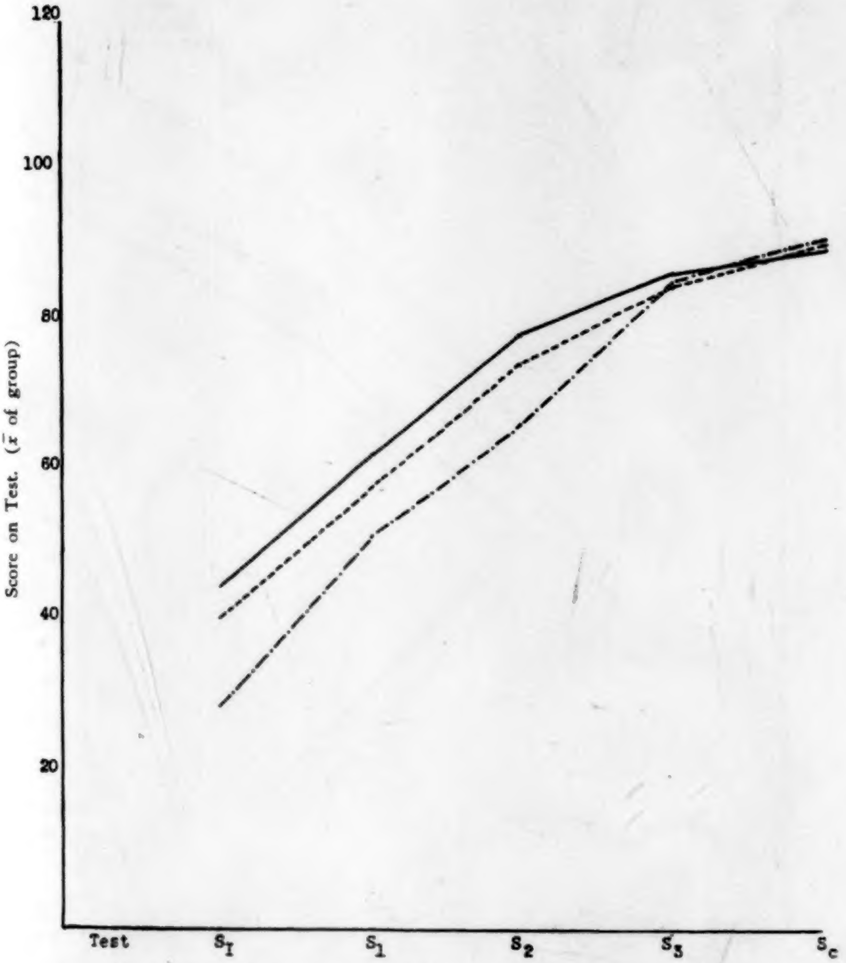

FIG. 7. BOYS AND GIRLS COMBINED. TESTER B. GRADE VIII

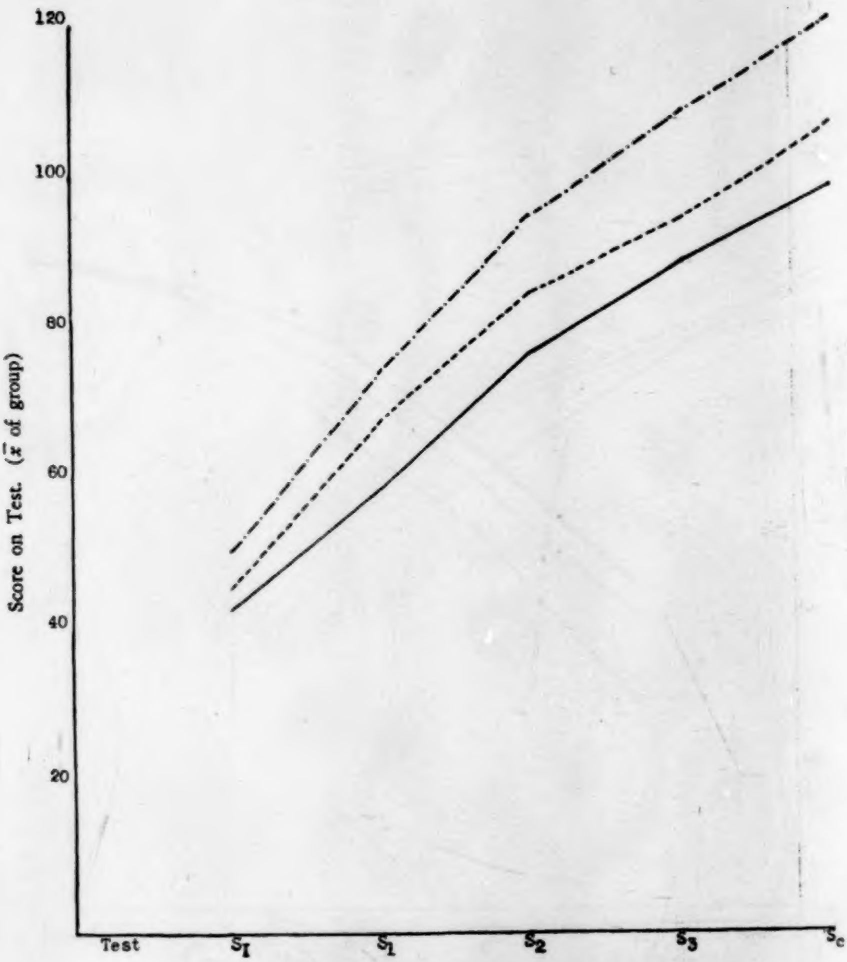

FIG. 8. ALL BOYS. TESTER B. HIGH SCHOOL

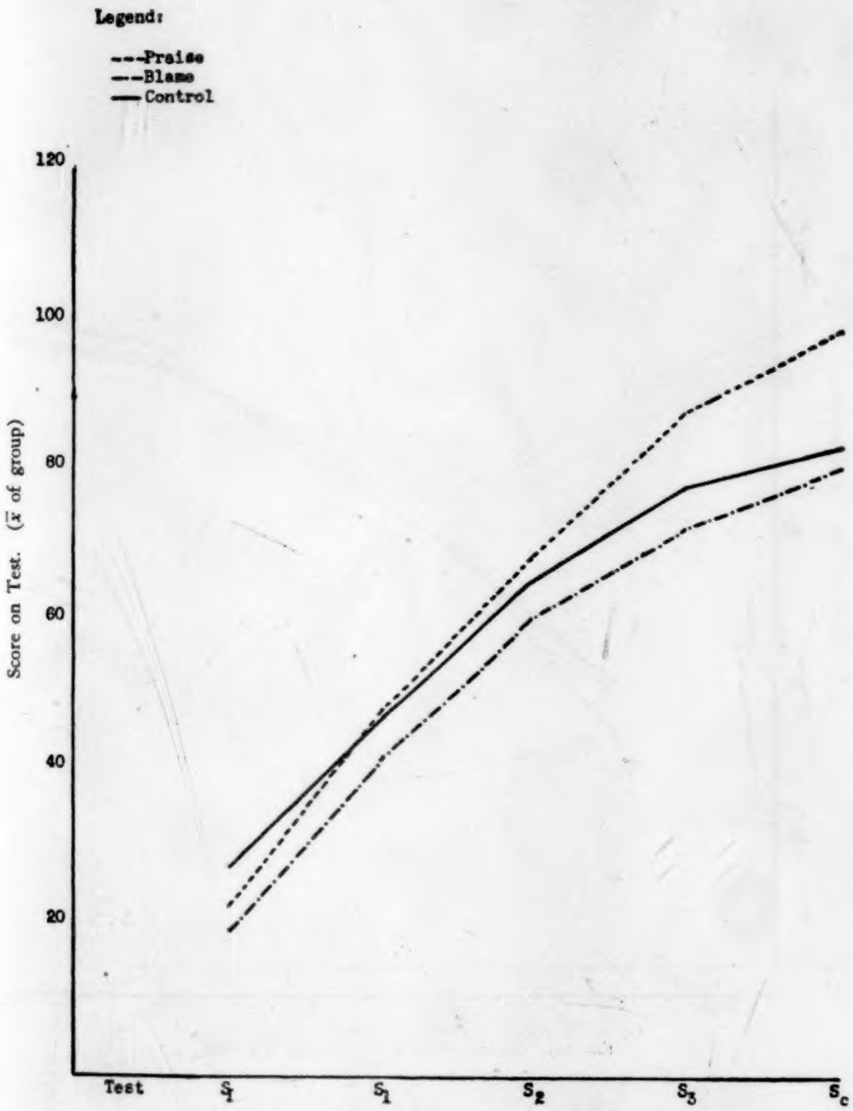

FIG. 9. BOYS. TESTER A. GRADE VII–VIII

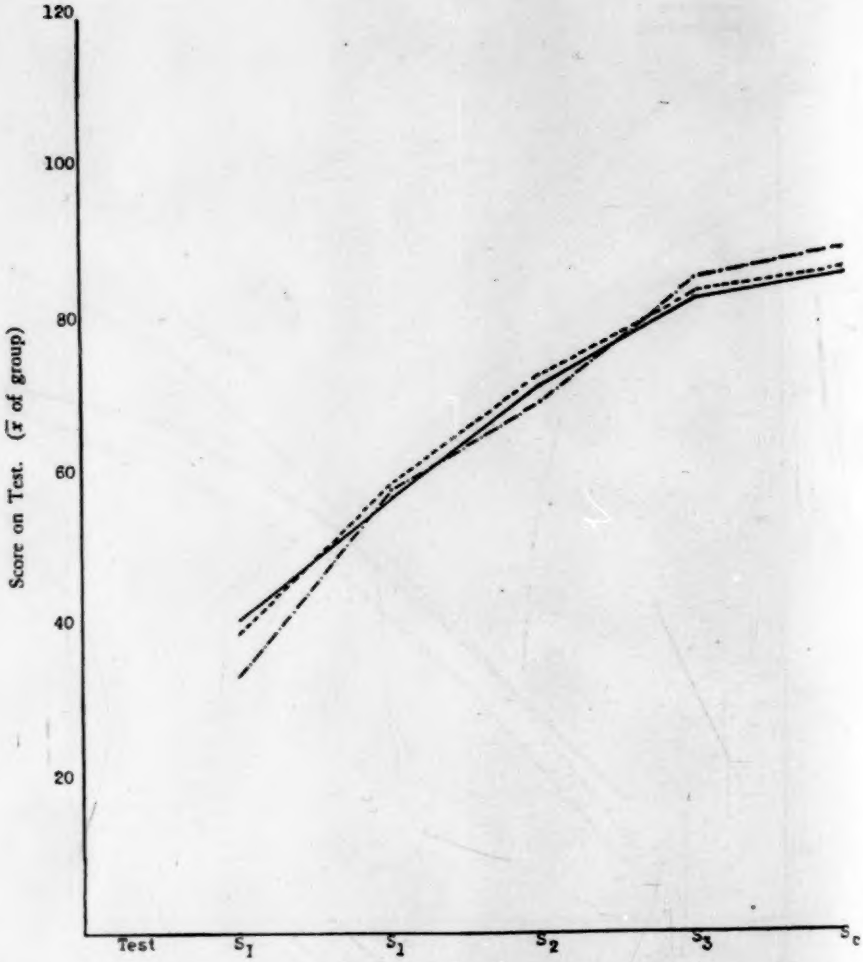

FIG. 10. GIRLS. TESTER A. GRADE VII–VIII

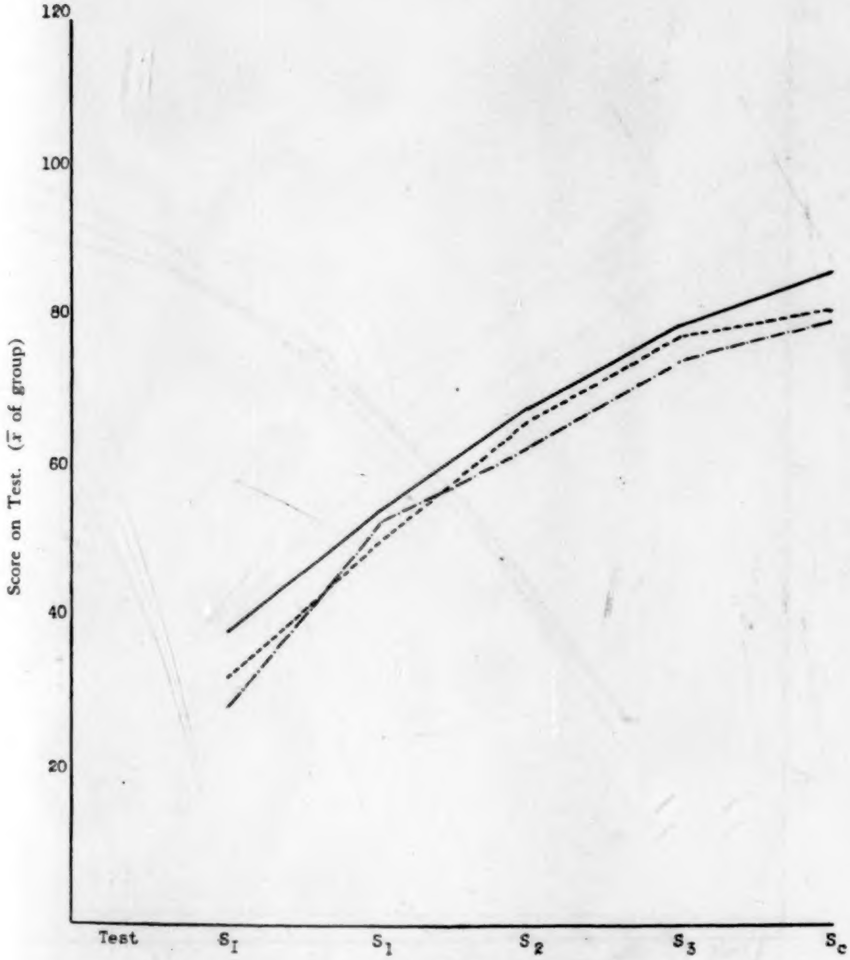

FIG. 11. BOYS. TESTER B. GRADE VII–VIII

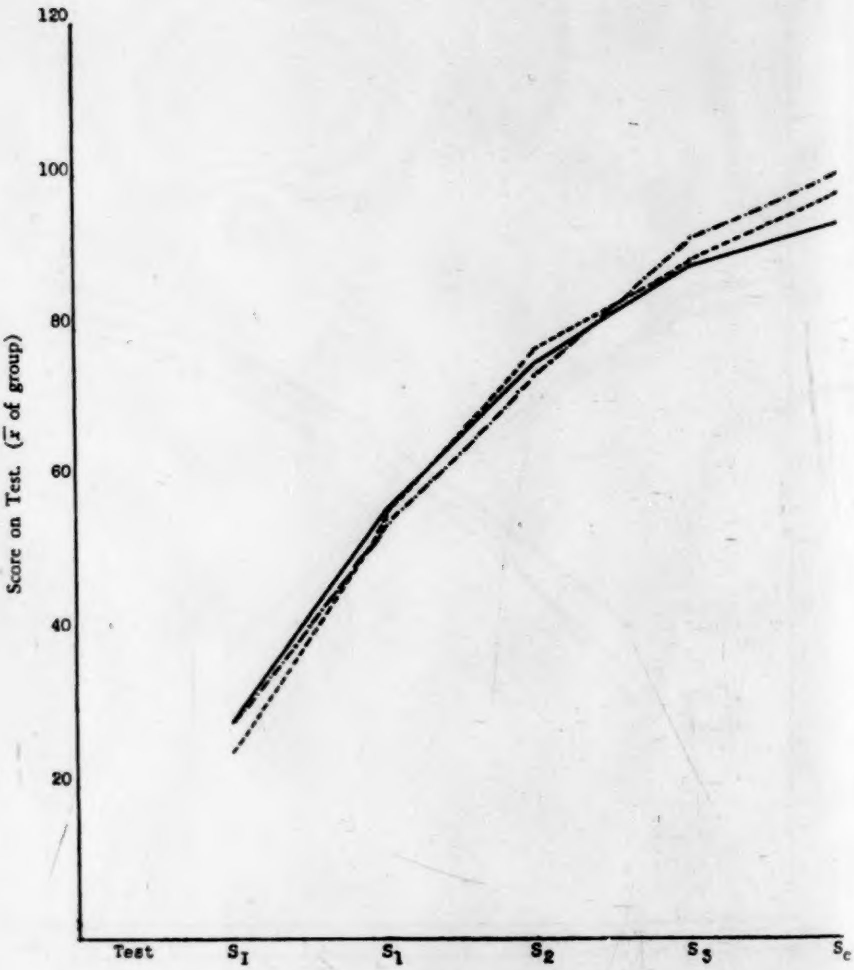

FIG. 12. GIRLS. TESTER B. GRADE VII–VIII

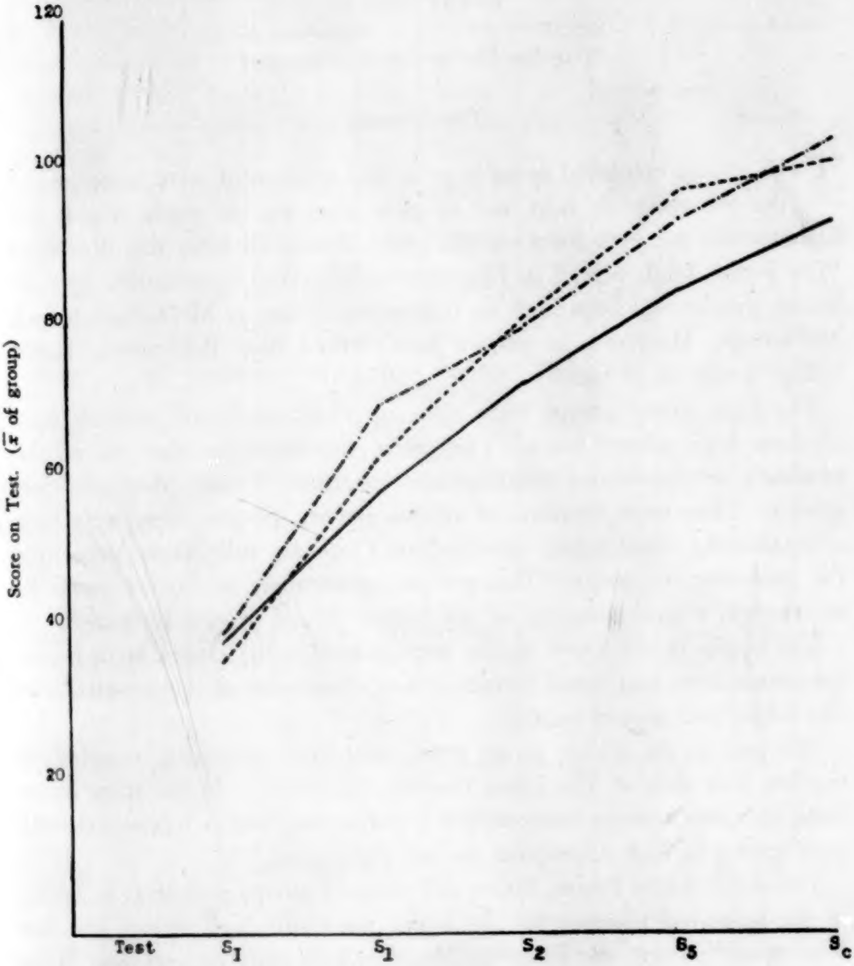

Fig. 13. All Men. Tester A. College

CHAPTER II

The Set-Up of the Experiment

The Groups

THE groups employed as subjects in this experiment were comprised of the following: 85 boys and 81 girls from six 7th grade classes, and 84 boys and 102 girls from six 8th grade classes, all from the Woodland Way Junior High School in Hagerstown, Maryland (population, approximately 35,000); 192 boys from six high school classes at McDonogh School, McDonogh, Maryland (a private boys' school near Baltimore); and a college group of 30 men.

The high school groups were made up from unselected students from all four high school years. They were unselected in that no regular scholastic or intellectual qualifications determined their placement into groups. They were members of various activity groups: some were from a tap-dancing class; others belonged to a chorus; still others were from the orchestra, et cetera. The groups represented, as far as could be ascertained, a true sampling of the Upper School of this institution.

The pupils in the lower grades were placed in the classes on a heterogeneous system, and could be taken as representative of the population of that school and general locality.

The men in the college group were, with three exceptions, members of the first year class at The Johns Hopkins University. In the three exceptions, one was a senior student, one a junior, and one a sophomore. All were men who had volunteered for the experiment.

The division into Praise, Blame and Control groups was done according to the following scheme: for the lower years and high school, the first class tested became the Praise group; the next class became the Blame group; and the last class made up the Control group. In the college group the men were arranged first according to intelligence test scores Otis (63); then into clusters of three, the first of this cluster becoming a Praise subject, the next a Blame subject, and the third, the Control subject.

The 7th and 8th grade groups and the high school groups were administered the tests as a group. The members of the college group were tested individually. In addition, a continuous kymogram was made on this latter group for pulse and respiratory rates, and for blood pressure.

TABLE I

SIMILARITY OF THE GROUPS WITH RESPECT TO CA, MA, AND INITIAL SCORE. TESTER A

		CA IN MONTHS			MA IN MONTHS			INITIAL SCORE		
GRADE	N	RANGE	$\bar{x}$	s	RANGE	$\bar{x}$	s	RANGE	$\bar{x}$	s
VII										
Praise	26	141–181	153.1	11.5	134–178	155.2	10.3	0–47	18.9	14.9
Blame	29	133–188	155.3	17.9	126–183	149.7	15.8	0–41	20.4	11.8
Control	30	142–177	153.1	12.1	134–174	148.1	9.5	8–40	21.6	8.9
VIII										
Praise	31	153–191	163.6	9.2	125–200	163.7	16.2	5–43	25.6	11.2
Blame	30	144–188	162.1	16.6	117–203	163.3	17.9	5–56	26.6	4.5
Control	33	143–180	159.9	8.9	133–201	162.3	16.0	7–64	33.3	14.9
VII–VIII										
Boys:										
Praise	29	141–181	159.0	12.2	125–193	158.2	15.7	0–45	21.2	14.8
Blame	29	133–188	155.2	14.4	117–203	155.7	20.9	0–41	19.3	12.1
Control	28	140–180	160.1	12.5	134–201	154.6	16.3	9–64	27.6	12.2
Girls:										
Praise	28	141–191	159.8	7.7	142–200	159.5	14.0	5–47	23.5	11.9
Blame	30	132–188	157.9	15.2	127–184	157.9	13.9	9–56	27.6	12.4
Control	35	142–178	155.9	9.2	133–196	156.3	12.4	7–52	27.6	15.1
H.S.										
Praise	14	175–215	190.7	12.9	160–242	202.4	25.0	5–52	32.9	11.2
Blame	41	169–220	196.2	13.5	146–263	196.2	26.2	4–57	32.9	11.2
Control	23	179–228	200.8	18.7	140–232	184.4	21.1	6–65	33.0	19.1
COLLEGE										
Praise	10	204–238	221.5	8.4	175–230	209.0	16.8	18–49	35.4	10.2
Blame	10	204–260	222.2	16.8	178–234	211.2	17.3	28–50	38.9	14.1
Control	10	203–258	224.8	15.0	193–230	210.9	12.5	27–69	37.6	12.2

TABLE II

SIMILARITY OF THE GROUPS WITH RESPECT TO CA, MA, AND INITIAL SCORE. TESTER B

		CA in Months			MA in Months			Initial Score		
Grade	N	Range	$\bar{x}$	s	Range	$\bar{x}$	s	Range	$\bar{x}$	s
VII										
Praise	29	130–188	156.7	17.6	126–176	147.1	12.7	3–63	30.8	13.9
Blame	30	137–181	152.7	11.4	126–183	154.2	13.2	7–55	34.3	12.2
Control	22	139–178	155.9	8.4	133–178	155.3	11.8	3–55	32.1	11.9
VIII										
Praise	31	152–196	165.4	11.4	131–197	159.4	15.3	22–68	40.5	10.9
Blame	28	146–185	164.3	10.4	130–190	159.9	16.8	5–61	29.3	14.3
Control	33	153–191	163.8	13.4	137–204	160.6	17.8	18–88	44.4	14.2
VII–VIII										
Boys:										
Praise	31	140–196	163.8	16.8	126–185	150.6	14.4	6–62	32.9	10.8
Blame	26	139–185	160.7	12.0	130–183	154.9	14.8	7–50	29.7	11.5
Control	26	143–186	162.9	5.6	125–204	154.8	16.7	18–88	37.9	14.8
Girls:										
Praise	29	130–196	158.5	14.8	135–197	156.5	14.7	3–68	38.8	15.2
Blame	32	137–183	153.8	12.3	126–190	158.4	15.6	5–61	33.6	14.6
Control	29	139–191	158.6	16.8	137–190	161.6	15.2	3–67	40.8	14.3
H.S.										
Praise	41	177–230	204.8	9.6	144–256	197.9	25.3	13–104	44.6	16.8
Blame	30	177–224	197.4	13.5	148–227	192.1	22.0	24–71	49.9	11.9
Control	43	168–223	196.7	10.9	157–241	197.8	12.4	14–69	41.7	6.7

All subjects received a code number, so that no names appeared on any of the test sheets.

The total number of subjects used was 574. Tables I and II show the divisions and similarities of the groups for chronological age, mental age, and initial score. Mental ages were computed from the intelligence quotients secured from the school, and in the case of the college group, from the results of the intelligence test given by the writer. Where the chronological age exceeded thirteen years, zero months, its multiplier was determined after Terman and Merrill's system (73).

The Selection of the Material

The test material was devised by the writer. It consists of a number of characters, six, from Gates' Test of Associative Learning (34),[1] arranged on a page. Figure 14 presents the key and first two lines (of 14) of the

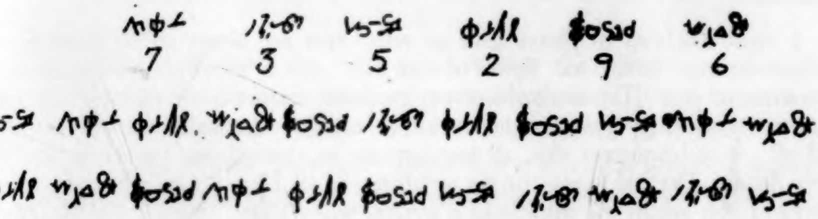

FIG. 14. KEY AND LINES 1 AND 2 OF TEST EMPLOYED

test. It is a code substitution test, the code appearing at the top of each sheet. The characters were first traced on a stencil. These were then mimeographed, cut out, and pasted on a large sheet of paper, the scaling being approximately 2:1 of the ordinary 8½ x 11-inch size paper. This large sheet was then reduced photographically: a new stencil tracing was made of this, and the requisite number of tests then were mimeographed.

The total number of characters that can be coded is 154. This is not quite sufficient number to give reasonable assuredness that no one will complete the test before the allotted time of two minutes. On several occasions it appeared that some subject would finish ahead of time on the last test. In these cases an additional sheet was given them just prior to the last test. Later analysis showed that this move had been well taken in several instances: one subject coded 157 characters, and another, 161.[2]

[1] The writer would like to take this opportunity to express his appreciation to Professor Gates for his kind permission to make use of this material.

[2] The original test contains 165 characters. In the tests used in this experiment one column was omitted because of the necessity of leaving a margin for the mimeograph machine. It had been planned to have the original test plated and printed, but the cost appeared prohibitive.

Each subject received a packet of five tests, each numbered serially and with subscripts in the upper right hand corner of the page. This would identify a subject and a particular test. Thus, a packet was numbered, for example, 1201_a, 1201_1, 1201_2, 1201_3, and 1201_c. The five tests were clipped together with a thin staple, and had an outside yellow "second-sheet" upon which also appeared the proper number and upon which the subject wrote his name at the termination of the test.

Procedure

Upon entering the room, the tester requested that all desks be cleared and then gave out the test packets. The subjects were cautioned beforehand, and this was seen to, that no one was to examine the packet nor in any other way get a "running start" on any of the others.

When all was in readiness the following instructions were read to the subjects:

I must ask you to please give to what you are about to do your most whole-hearted attention. Some of this may appear as silly, or monotonous, to some of you. Try not to have any personal views on the matter. Do not worry about it if you cannot do all or many: practically no one can do it all. It is important that all instructions be carried out promptly and to the letter. Do not begin, or do anything, until I say "Go"; stop at once, even if you are in the middle of a stroke, when I say "Stop." Do not turn any pages until told to do so.

This is a code, or substitution, test. The object is to see how many of the numbers you can place correctly *underneath* the proper symbols. The key is at the top of each sheet. Work from left to right in each row. When I give the signal to begin, turn the first sheet (the topmost sheet as the papers are in front of you) and start working at once. No questions will be answered after the test begins. *(Pause.)* Everybody ready? Begin.

The examiner walked about while the subjects were working, observing their work and making certain that all were on the correct page. At the end of exactly two minutes, the examiner said:

Stop. Everybody stop. Tear off the sheet upon which you have just been working and pass the sheets forward quickly.

Care and insistence had to be exercised at this point in order to prevent counting-up of scores on the part of the subjects, or discussion with their neighbors.

After the papers had been collected the tester made an apparent intensive examination of the papers for one minute, keeping alert, nevertheless, to prevent talking, or the turning over to the next sheet.

At the end of the minute interval, the examiner said (depending upon the incentive to be employed): "That was very good (poor)," or made no comment. "Let's try this again. Begin when I say 'Go'; stop when I say 'Stop.' Ready, begin." The procedure of walking about, collecting and checking papers, etc., was repeated. Before each of S_2 and S_3 the tester said: "You are doing very well (poorly)," or made no comment. "Let's try this again. Ready, begin."

As the preface to S_e, the tester announced the following: "You have done very well (poorly)," or made no comment. "This is the last test. Now, I shall give one dollar to the one showing the greatest improvement. Ready, begin." At the completion of this test, the papers were passed forward as before; then the pupils were asked to write their names on the yellow facing sheets, which were collected.

Each separate test period was exactly two minutes in duration, and each interval between tests one minute. Timing was done with stop watches which were checked with each other and found to have no detectable differences.

There were two testers for each grade, except the college group. The two testers worked independently of each other, but at the same time and at the same school. These testers will be known as Tester A and Tester B. Tester A was the writer; Tester B was a female graduate student in Education at Johns Hopkins, with some training in test procedure, and who had been thoroughly instructed and rehearsed by the writer. Each tester carried a set of directions and procedure into each class and followed these implicitly.

As a check against confusion in spite of coding, each separate test was immediately bound with an elastic band after the one minute examination period and labeled. The completed tests of each class were kept in separate manila folders which were marked with the class designation, the incentive employed, the place, and the tester.

The College Group

The college group was tested by Tester A only. These were all individual tests. Before the test proper began, each subject was questioned as to the amount of exertion undergone, or as to the amount and time of food taken, immediately prior to appearing for testing. Nothing was done in the case of recent partaking of food (there were three cases); but in the case of exertion, a reasonable time was allowed for recovery. After this, the

apparatus for recording continuous blood pressure, pulse, and breathing rates was adjusted and set into operation. There was a preliminary (S_p) period of two minutes duration with each subject, during which they performed some simple task, such as Gates'[3] Same-Different Digits or Same-Different Figures. After an interval of one minute, the test proper began and followed the same procedure as described earlier. At the end of the test proper there was a two-minute "final" phase (S_F), during which the subject rested and engaged in small talk with the writer.

Care was observed that the conservational topics were quite banal and that, in so far as could be judged, they were non-stimulating. The sole purpose of the conversation was to avoid ennui on the part of an otherwise unoccupied subject.

Each subject had been advised before the beginning of the experiment that there would be no "tricks," or shocks, or fright-situations; and they were adjured to remain as motionless as possible during the running of the experiment. They were advised, too, that possibly unusual sensations of tingling or coldness might be felt in the arm or hand employed with the sphygmomanometer cuff, but were asked to disregard these. In no case did a subject report annoyance or discomfort from the apparatus. In fact, all stated that they were unaware of any sensation in the arm or hand, or chest, until after termination of the experiment. The amount of discomfort they reported as being of negligible degree.

All were made as comfortable as possible before beginning the experiment. Each group reported at approximately the same time of day. With four exceptions, the members of a cluster followed each other at about twenty-four hour intervals. In the case of the four exceptions, there was a ten-day lapse. These men had not reported when first scheduled because of conflicts in classes, and reported after the Christmas holiday.

The Apparatus

The apparatus employed with the college subjects was standard laboratory equipment for measuring the physiological changes of blood pressure, pulse, and breathing rate. The subject was in every instance screened from the apparatus. The apparatus consisted of a recording drum, with extension, revolving at the rate of 3.25 inches per minute; a sphygmomanometer of the common desk type, with a Schrader air release valve between the top of the manometer and tambour cut-off valve, tambour and recording lever;

[3] *Op. cit.*

Sumner-type pneumograph, with tambour and recording lever; signal marker, operated from a single dry cell in series with a Jacquet chronometer registering at one second intervals, and a manually operated signal marker, operating off the same dry cell.

Thoracic breathing curves were obtained, the rubber tubing being placed in each case across the sterno-costal angle anteriorly, and the chain passing just below the inferior margins of the scapulae posteriorly. Tension was approximately the same for each subject. In addition, in order to further insure comfort to the subject, the plates of the pneumograph were placed at the mid-axillary line.

Since the average time each subject spent in the apparatus was approximately seventeen and one-half minutes, no large amount of pressure could be placed against the circulation. Preliminary experimentation had shown that at about eighty millimeters pressure of mercury, the optimal comfort to the subject and identifiable pulse readings could be obtained. This is, of course, neither a systolic nor a diastolic pressure; but variation in blood pressure, whatever it may be, can be obtained, as well as pulse beats, with this arrangement.

Scoring

All scoring was done by the writer. No objective attempt was made to determine whether or not the subjects' responses were correct. His score was the number of substitutions attempted. This was considered a legitimate procedure since the product-moment correlation, based on a sampling of 100 papers, between number of items attempted and number of correct responses, was +.944.

In the case of pulse and breathing rate, only the rates during a period thirty seconds before the end of a test and thirty seconds after its beginning were used. The interval between tests was used in its entirety. With the breathing rate the I-factor, or I/E ratio, was computed; while for pulse, an average rate per minute was computed from the rates of the sample periods.

CHAPTER III

The Results *

General Treatment

THROUGHOUT the discussion, and in the various tables, the notations $\bar{x}$ and s will be used respectively as estimates of m, the mean of the population, and σ, its standard deviation. Where N was small, as it was in most of the cases, σ was best estimated from the following expression (*Goulden*, 37):

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}$$

The difference in score of the last test over the first test, or gain, was employed as the measure of learning. The Pearson Product-Moment correlation of Gains vs. Final Score was in every instance positive and high (Tables III and IV).

Before the significance of gains and/or their relatedness could be analyzed, there were several variable factors, or possible influencing elements, that demanded *a priori* consideration. These were: chronological age, mental age, initial score, and objective features pertaining to the two testers. There conceivably might be added the matter of *locale*, or general setting in which the testing was conducted.

What could be done about eliminating these influences, or about holding them constant, presented a real although no novel problem. The question of equating one's subjects or groups is one with which every experimenter is faced in a problem of this sort. Actual physical equating of subjects by elimination presents several disadvantages; one of these is that often the numbers with which one is left have been reduced to a point beyond which even small sampling technique would prove untrustworthy; another, and perhaps even more vital point, is the somewhat intangible one of just what happens to data, or just what sort of data are left, if cases have been extracted from them.

* The scatter-diagram worksheets for the correlations, and original data sheets, are not presented but are on file at The Johns Hopkins University.

TABLE III

PEARSON PRODUCT-MOMENT CORRELATIONS (INCENTIVE GROUPS AND SEXES COMBINED). GRADES VII, VIII, H.S., AND COLLEGE

	VII		VIII		H.S.		COLLEGE
	n,85	n,81	n,94	n,92	n,78	n,114	n,30
GROUPS	A	B	A	B	A	B	A
Gains vs. CA	.064	.245	−.156	.110	−.227	−.213	−.336
Gains vs. MA	.176	−.227	.087	.112	.285	.077	.166
Gains vs. IS	−.254	−.135	−.079	−.237	−.041	−.088	.260
Gains vs. FS	.863	.831	.855	.760	.881	.858	.811
IS vs. FS	.358	.418	.447	.355	.539	.428	.499
MA vs. IS	.263	.380	.180	.180	.149	.198	.207
MA vs. FS	.301	.002	.259	.185	.303	.143	.231
CA vs. IS	−.162	−.210	−.132	−.123	−.177	.137	.151
CA vs. FS	−.112	.195	−.239	.017	−.243	−.139	−.205
MA vs. CA	−.298	−.491	−.374	−.406	−.360	−.226	−.199

As to the matter of *locale,* it was not considered necessary or feasible to attempt corrections for S.E.S., or similar influences, since our groups were drawn at random and were representative of the population from which drawn. It is conceivably true, of course, that there may be differences between major groups.

The extrinsic or objective influence of each tester was held constant in so far as was possible to contemplate beforehand; *e.g.,* each tester wore his or her same outer garments at each testing; reading of directions was rehearsed, so that no voice or other cues would be present in one situation and not in another; and procedure was *standardized* as given in Chapter II.

In this study the difficulties usually encountered in equating groups have been overcome through the employment of a statistical procedure. The

TABLE IV

PEARSON PRODUCT-MOMENT CORRELATIONS (INCENTIVE GROUPS COMBINED). GRADES VII–VIII COMBINED

	A		B	
GROUPS	n,86	n,93	n,83	n,90
	B	G	B	G
Gains vs. CA	−.09	−.38	.25	.17
Gains vs. MA	.06	.26	.11	.22
Gains vs. IS	−.12	−.10	−.19	−.23
MA vs. IS	.45	.38	.23	.32
CA vs. IS	−.38	.10	−.07	−.13
MA vs. CA	−.28	−.10	−.34	−.23

number of cases when broken into various groups, although not large, were of sufficient size to permit the use of small sampling techniques, and the data were left in the original unadulterated state. The variables that were accounted for or corrected for by the method adopted were those of chronological age, mental age, and initial score.

However, before the data could be treated, it was first necessary to test the linearity and normality of the various distributions. This was done by plotting the regressions of Gains vs. MA, Gains vs. CA, etc. These curves were found to be sufficiently close to normal and linearity to be best fitted by a straight line, and to permit the calculation of multiple regression equations (Holzinger, 40).

Throughout this study where comparisons have been made, or the symbol "vs." employed, it is to be understood that the difference indicated is that of the first member *minus* the second member. Hence, there will appear at times negative differences; and, in many cases where variable-correction factors have been applied, it will be found that the corrected difference is larger than the raw difference.

In making comparisons, use has been made of the *relative deviate,* which is merely the number of standard deviations that x, a variate, is away from

the mean of its series, in either direction. This can be expressed thus (Treloar, 76):

$$k = \frac{x - \bar{x}}{s_x}$$

where x-$\bar{x}$ and s_x are dimensions on the x scale. Dividing one by the other produces k, a pure number, the original units of measurement (age, IQ, etc.) vanishing. From this one can secure P values from a table of normal curve functions, P being the probability of k being exceeded through errors of random sampling.

Expanding the preceding equation to fit the data of this experiment, there was used, as typical, this form:

$$k = \frac{(\bar{X}_B - \bar{X}_c) - b_{\overline{x \cdot yzw}}(\bar{Y}_B - \bar{Y}_c) - b_{\overline{x \cdot zyw}}(\bar{Z}_B - \bar{Z}_c) - b_{\overline{x \cdot wyz}}(\bar{W}_B - \bar{W}_c)}{x \cdot yzw \sqrt{\frac{1}{n_B} + \frac{1}{n_c}}}$$

where k is the relative deviate for a difference in gain of a Blame group versus a Control group, the difference being corrected by multiple regression correlations for the variables of chronological age (Y), mental age (Z), and initial score (W), respectively. The constants employed were found from product-moment correlations which are shown in Tables III and IV, pages 31 and 32.

In testing the significance of a difference between two means, the *level of significance* chosen was $k = 2.00$, $P = .0455$. That is, if P were larger than .0455, the hypothesis that the difference was a real difference and not due to errors of random sampling was rejected.

Examination of the Learning Curves

Word picturization of curves are very apt to be unsatisfactory and poor substitutes for the curves themselves. Consequently, they will be dealt with rather summarily, the attempt being made only to designate some of the most salient features. Tables V and VI present the scores made by the various groups on the several tests. Figures 3 through 13 are the graphs for these scores. From examination of the curves several things will be noted: (1) There appears no markedly age or grade differentiation in the general slope of these curves. With advancing status, the initial point is higher, but so is the point reached on S_c. The curves are flatter for Tester B than for Tester A. (2) In no instance does a curve fall below a point previously reached. Even where a loss in acceleration in the curve appears, the lesser improved score is still above that of the preceding test. This may be accounted for, in part at least, to practice effect. That is, even though the specific incentive, or some other extrinsic or intrinsic factor, may have

TABLE V

SCORES ($\bar{x}$ and s) ON THE FIVE TESTS. TESTER A

		S_I		S_1		S_2		S_3		S_C	
GRADE	N	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
VII											
Praise	26	18.9	14.9	49.9	22.3	69.9	23.1	85.2	23.4	97.3	24.9
Blame	29	20.4	11.8	42.3	16.7	62.5	17.5	73.9	28.6	81.3	24.3
Control	30	21.6	8.9	42.7	15.0	58.1	21.1	72.2	24.6	79.9	25.7
VIII											
Praise	31	25.6	11.2	54.7	17.4	73.8	18.5	89.3	21.0	98.3	22.6
Blame	30	26.6	4.5	54.2	16.2	70.2	22.5	89.6	26.6	97.7	29.2
Control	33	33.3	14.9	60.9	22.3	80.4	26.4	93.3	24.6	97.2	27.1
VII–VIII											
Boys:											
Praise	29	21.2	14.8	47.8	19.6	68.6	21.3	87.2	22.9	98.5	21.9
Blame	29	19.3	12.1	41.1	16.3	60.3	20.8	72.6	23.4	80.2	25.8
Control	28	27.6	12.2	47.4	19.9	65.1	26.1	77.7	27.9	83.3	25.3
Girls:											
Praise	28	23.5	11.9	55.5	16.0	76.0	18.9	87.7	21.4	97.1	25.0
Blame	30	27.6	12.4	55.0	16.4	72.3	18.6	90.8	24.2	99.2	26.9
Control	35	27.6	15.1	55.8	22.2	74.2	25.5	87.4	25.6	93.4	29.0
H.S.											
Praise	14	32.9	11.2	60.6	16.4	78.5	21.5	91.1	23.4	104.0	27.5
Blame	41	32.9	11.2	59.9	19.1	82.9	19.9	97.8	21.4	107.4	22.9
Control	23	33.0	19.1	57.8	18.8	74.2	23.7	87.9	22.8	98.9	27.3
COLLEGE											
Praise	10	35.4	10.2	61.1	14.2	80.4	19.1	96.8	23.2	100.1	17.5
Blame	10	38.9	14.1	69.3	12.7	79.6	13.3	92.9	17.6	103.3	22.2
Control	10	37.6	12.2	56.5	12.8	71.8	18.9	83.4	28.7	91.0	25.0

TABLE VI

SCORES ($\bar{x}$ and s) ON THE FIVE TESTS. TESTER B

		S_I		S_1		S_2		S_3		S_c	
GRADE	N	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
VII											
Praise	29	30.8	13.9	50.6	21.2	66.1	25.2	74.0	24.9	78.1	26.6
Blame	30	34.3	12.2	54.0	18.6	68.0	23.0	75.9	18.1	83.1	23.0
Control	22	32.1	11.9	48.3	16.9	60.4	17.1	73.8	12.3	83.1	19.1
VIII											
Praise	31	40.5	10.9	57.9	18.7	73.0	19.3	83.8	17.9	88.6	39.2
Blame	28	29.3	14.3	51.1	18.7	65.2	20.9	84.0	19.8	88.8	21.2
Control	33	44.4	14.2	61.5	22.8	76.9	20.5	85.7	21.1	88.4	23.5
VII–VIII											
Boys:											
Praise	31	32.9	10.8	50.6	15.4	66.8	19.8	77.1	19.2	81.9	22.9
Blame	26	29.7	11.5	52.3	14.9	63.0	16.9	74.7	14.8	81.2	17.8
Control	26	37.9	14.8	54.4	22.8	68.1	23.4	78.8	20.9	86.3	24.2
Girls:											
Praise	29	38.8	15.2	58.4	21.3	72.7	24.4	83.3	15.7	86.4	24.9
Blame	32	33.6	14.6	58.3	20.5	69.7	25.1	85.0	15.1	89.8	24.8
Control	29	40.8	14.3	57.8	20.5	71.2	21.3	82.9	19.1	86.2	19.9
H.S.											
Praise	41	44.6	16.8	67.0	24.9	84.2	24.6	94.5	28.2	106.0	27.4
Blame	30	49.9	11.9	74.2	19.8	94.0	19.5	107.8	22.9	120.9	21.9
Control	43	41.7	6.7	57.9	17.7	76.3	21.9	89.3	17.7	98.3	23.6

been operating to inhibit further progress, they were not of sufficient potency to overcome amount of progress already made. (3) There is a distinct tendency for the rate of improvement to fall off between S_3 and S_c. This

is evident in all groups in Figures 4, 7, and 12; while it is manifest for Blame and for Control groups in Figure 3; for Praise, in Figure 13; and for Control, in Figure 9. This may have been brought about through the common incentive that was applied having an over-stimulating effect. That is, the subjects were *so anxious* to succeed that the feeling tone aroused actually produced inhibitory action. (4) In no other respects are the curves found to have consistently common characteristics. For example, in Figure 3: Control group leads at S_I, holds this over Blame only at S_1, then falls below both of the other groups from there through S_e; whereas, the Praise group starts with the lowest score, jumps definitely into the lead at S_1, and continues to increase this lead to the end. In Figure 13, Blame assumes an early lead, loses it to the Praise group, then leads both Praise and Control at S_e. In Figure 11, Control is superior to the other groups throughout all the tests.

In other cases, groups, as in Figure 7, which show little effect from the specific incentives are almost coincident at S_e; while in still other cases, as in Figure 10, groups which show some effect of the incentives show a larger effect at the end.

In yet other instances, as in Figure 4, groups which are in the lead to S_3 fall below at S_e; while groups, as in Figure 6, which have trailed the others right along assume the lead on the final test.

In regarding Figures 9, 10, 11, and 12, one is struck by the appearance of the curves. Those for the boys, with each tester, follow more or less parallel courses, while those for the girls are quite close together and interwoven. It would appear from this on first analysis that the boys are more affected than the girls by the test situation in general, than the specific stimuli, in particular. That is, it would seem that the preliminary "set," or the "goal-set," of the boys was more evident, more diversified, and less subject to extrinsic motivation than was the case with the girls. The girls' "set" was more common to all the groups, but was the more readily affected by extrinsic factors. However, the closeness of the curves to each other for the girls may be giving to their interweaving exaggerated impressions of their variability, or affectivity.

Discussion of the S_e-S_I Gains

Grades

Tables VII, VIII and IX present the raw and corrected gains (S_e-S_I scores) of the various groups arranged according to grade. Table VIII shows the significance of the differences for the gains uncorrected. Table IX shows the significances corrected for the influence of MA, CA, and

Initial Score for grades VII and VIII, high school and college. From Table IX, it will be observed that in grade VII with the one tester the Praise group shows a significant gain over both the Blame and Control groups: P=.0050 and .0009, respectively. With the other tester no gains are of statistical significance.

TABLE VII

RAW GAINS ($\bar{x}$ and s)

	PRAISE			BLAME			CONTROL		
GRADE	N	$\bar{x}$	s	N	$\bar{x}$	s	N	$\bar{x}$	s
VII									
Tester A	26	78.4	22.3	29	60.9	22.4	30	58.3	23.9
Tester B	29	47.3	25.6	30	48.8	23.1	22	51.0	12.5
VIII									
Tester A	31	72.7	20.9	30	71.1	26.2	33	63.9	23.7
Tester B	31	48.1	18.3	28	59.5	21.7	33	44.0	19.3
H.S.									
Tester A	14	71.1	21.4	41	74.5	21.5	23	65.9	18.7
Tester B	41	61.4	26.3	30	71.0	20.1	43	56.6	21.2
COLLEGE									
Tester A	10	64.7	19.3	10	64.4	19.9	10	53.4	25.2

In grade VIII, the gains of the Blame group are significantly different with Tester B: P=.0000 for the Blame group over the Control, and P=.0002 for Blame over Praise. With the other tester, none of the differences is of significance.

At the high school level, only Blame over Control with Tester B bears statistical significance: P=.0060.

With the college group none of the differences is significant.

Sexes

In the foregoing portrayals of the curves and scores of the groups divided as to grades, both boys and girls were included in grades VII and VIII. Because of the smallness of numbers, it was not considered feasible to

separate the sexes as to grade. Instead, the sexes were separated by grouping together both seventh and eighth grades (designated as grade VII-VIII combined). Tables I and II (pp. 23 and 24) show the similarities of the two years when separated and when combined with respect to MA, CA, and Initial Score.

TABLE VIII

SIGNIFICANCE OF DIFFERENCE IN GAINS, UNCORRECTED FOR INFLUENCE OF CA, MA, AND IS. GRADES VII, VIII, H.S. AND COLLEGE

	Tester A			Tester B		
Grades	Diff.	k	P	Diff.	k	P
VII						
P vs. C	20.1	3.26	.0011	−3.7	0.68	.4965
B vs. C	2.6	0.45	.6527	−2.2	0.44	.6599
P vs. B	17.5	2.88	.0040	−1.5	0.24	.8103
VIII						
P vs. C	8.8	1.60	.1096	4.1	0.86	.3898
B vs. C	7.2	1.13	.2585	15.5	2.92	.0035
P vs. B	1.6	0.27	.7872	−11.4	2.17	.0300
H.S.						
P vs. C	5.2	0.75	.4533	4.8	0.92	.3576
B vs. C	8.6	1.68	.0930	14.4	2.96	.0031
P vs. B	−3.4	0.52	.6031	−9.6	1.75	.0801
COLLEGE						
P vs. C	11.3	1.13	.2585			
B vs. C	11.0	1.08	.2801			
P vs. B	0.3	0.03	.9761			

Tables X, XI, and XII show the raw gains and significance of the differences in scores for grade VII–VIII combined with the subjects separated as to sex. In Table XI, it will be noticed that only in the case of boys with Tester A is Praise significant over both Control and Blame:

P=.0000 and .0042, respectively; while only Blame over Control shows a significant difference with the girls for Tester B: P=.0006.

In Table XII, the differences in corrected gains are matched as to the

TABLE IX

SIGNIFICANCE OF DIFFERENCE IN GAINS, CORRECTED FOR INFLUENCE OF CA, MA, AND IS. GRADES VII, VIII, H.S. AND COLLEGE

	Tester A			Tester B		
Grades	Diff.	k	P	Diff.	k	P
VII						
P vs. C	19.9	3.31	.0009	−2.20	0.42	.6745
B vs. C	1.5	0.26	.7949	−1.40	0.29	.7718
P vs. B	16.9	2.81	.0050	−0.60	0.12	.9045
VIII						
P vs. C	8.74	1.48	.1389	5.80	1.23	.2187
B vs. C	7.50	1.25	.2113	23.10	4.84	.0000
P vs. B	1.34	0.22	.8259	−17.00	3.73	.0002
H.S.						
P vs. C	3.70	0.55	.5823	4.20	0.64	.5222
B vs. C	7.40	1.48	.1389	13.10	2.75	.0060
P vs. B	−3.44	0.60	.5484	−7.20	1.28	.2005
COLLEGE						
P vs. C	15.20	1.72	.0854			
B vs. C	11.60	1.32	.1868			
P vs. B	3.60	0.40	.6892			

sexes for the combined grade VII–VIII. In only one instance, that of Control girls over Control boys, is there apparent a significant difference between the sexes: P=.0285.

The general lack of statistically reliable differences in gains between the sexes may be a partial answer at least to the observation made on page 36 of the apparent wide dissimilarity in the learning curves for boys and girls.

The Different Testers

If the various initial scores be recalled, or reference made to Tables I and II and the various graphs, it will be observed that in every instance

TABLE X

RAW GAINS. BOYS AND GIRLS. TESTER A AND TESTER B DIFFERENTIATED. GRADES VII–VIII COMBINED

	Praise		Blame		Control	
Tester	B	G	B	G	B	G
A	77.3	73.6	60.9	71.6	55.7	65.8
n,	(29)	(28)	(29)	(30)	(28)	(35)
B	49.0	47.6	51.5	56.2	48.4	45.4
n,	(32)	(29)	(26)	(32)	(26)	(29)

TABLE XI

SIGNIFICANCE OF DIFFERENCE IN GAINS, CORRECTED FOR INFLUENCE OF CA, MA, AND IS. BOYS AND GIRLS. GRADES VII–VIII COMBINED

			Tester A					Tester B		
Groups	N		Diff.	k	P	N		Diff.	k	P
Boys										
P vs. C	29	28	26.8	1.35	.0000	32	26	3.30	0.66	.5093
B vs. C	29	28	9.0	1.47	.1416	26	26	6.80	1.34	.1802
P vs. B	29	29	17.7	2.86	.0042	31	26	−3.50	0.68	.4965
Girls										
P vs. C	28	35	1.54	0.41	.6818	29	29	7.00	1.23	.2137
B vs. C	30	35	3.41	0.93	.3524	32	29	19.40	3.44	.0006
P vs. B	28	30	−0.65	0.17	.8650	29	32	−6.20	1.10	.2713

the subjects made a larger initial score (averaging approximately ten points at each level) with Tester B than with Tester A. Tables XIII through XVIII show the various differences in raw and corrected scores and gains between the subjects of Tester A and Tester B. Table XVIII is a composite table of gains (raw scores).

Examination of Table XIII will show that in every instance but two the initial scores of Tester B are significantly higher than those of Tester A. The two cases that are exceptions are in grade VIII, Blame, and grade VII–VIII, girls, Blame, where P=.3371 and .1010, respectively.

It can be seen from Table XV that the subjects of Tester A gained substantially and significantly over those of Tester B. In only two instances could the difference be attributable to errors of random sampling. In these two cases, P was equal to .3271 for high school, Blame, and equal to .0488 for high school, Control. In this latter case the null hypothesis approaches very closely to being rejected.

TABLE XII

SIGNIFICANCE OF DIFFERENCE OF GAINS, CORRECTED FOR INFLUENCE OF CA, MA, AND IS. BOYS VS. GIRLS. GRADES VII–VIII COMBINED

	Tester A					Tester B				
Groups	N					N				
	B	G	Diff.	k	P	B	G	Diff.	k	P
P vs. P	29	28	6.9	1.16	.2460	31	29	3.4	0.67	.5029
B vs. B	29	30	-9.0	1.50	.1336	26	32	-3.4	0.66	.5093
C vs. C	28	35	-12.5	2.19	.0285	26	29	4.1	0.76	.4473

The effect of the correction may be observed where Tester A and Tester B are compared. Looking at Table XIV, uncorrected gains, it is observed that in four instances out of the nine significant differences are apparent. In Table XV, where the gains have been corrected for the effect of MA, CA, and Initial Score, seven of the nine differences are now of significance. In all instances the differences were in favor of Tester A.

Differentiating the sexes, grades VII–VIII combined, it will be observed from Table XVII that again significant differences exist between testers, with the increase on the side of Tester A. Here, in only one case, Control boys, can the difference be ascribed to chance: P=.0819. Yet this is quite close to being acceptable and is certainly in the trend with the other groups.

The question arises as to the possible causes of the difference in gains in favor of Tester A over Tester B. Is this difference attributable to the lower scores on S_I for Tester A, thus making the difference in gains a function of S_I, or are the scores on S_c higher for Tester A than for Tester B?

TABLE XIII

SIGNIFICANCE OF THE DIFFERENCE BETWEEN TESTER A AND TESTER B ON THE FIVE TESTS. (RAW SCORES)

	N		S_I			S_1			S_2			S_3			S_C		
	A	B	DIFF.	k	P	DIFF.	k	P	DIFF.	k	P	DIFF.	k	P	DIFF.	k	P
VII																	
Praise	26	29	−11.9	3.08	.0021	−0.7	0.12	.9045	3.8	0.67	.5029	11.2	1.72	.0854	19.2	2.59	.0090
Blame	29	30	−13.9	4.50	.0000	−11.7	2.54	.0111	−5.5	1.04	.2983	−2.0	0.38	.7029	−1.8	0.31	.7566
Control	30	22	−10.5	3.40	.0007	−5.6	1.24	.2150	−2.3	0.43	.6672	−1.6	0.31	.7566	−3.2	0.58	.5619
VIII																	
Praise	31	31	−14.9	5.30	.0000	−3.2	0.68	.4965	0.8	0.17	.8650	5.5	1.12	.2627	9.7	1.26	.2077
Blame	30	28	−2.7	0.96	.3371	3.1	0.67	.5029	5.0	0.88	.3789	5.6	0.92	.3576	8.9	1.33	.1835
Control	33	33	−11.1	3.17	.0015	−0.6	0.11	.9124	3.5	0.60	.5485	7.6	1.32	.1868	8.8	1.42	.1556
VII–VIII																	
Boys:																	
Praise	29	31	−11.7	3.73	.0002	−2.7	0.60	.5485	−1.9	0.36	.7188	10.1	1.83	.0672	16.6	2.86	.0042
Blame	29	26	−10.4	3.32	.0009	−11.2	2.67	.0076	−2.7	0.53	.5961	−2.0	0.41	.6818	−1.0	0.17	.8650
Control	28	26	−10.3	2.82	.0048	−7.0	1.19	.2340	−3.0	0.45	.6527	−1.0	0.15	.8808	−3.0	0.43	.6672
Girls:																	
Praise	28	29	−15.3	4.25	.0000	−2.9	0.59	.5552	3.3	0.57	.5687	4.4	0.71	.4777	10.7	1.62	.1052
Blame	30	32	−6.0	1.64	.1010	−3.3	0.70	.4839	2.6	0.46	.6455	6.0	1.17	.2420	9.4	1.43	.1527
Control	35	29	−13.2	3.57	.0004	−1.9	0.36	.7188	3.0	0.51	.6101	4.6	0.32	.4122	7.3	1.16	.2380
H.S.																	
Praise	14	41	−11.7	3.00	.0027	−6.6	1.12	.2627	−5.7	0.83	.4065	−3.4	0.44	.6599	−2.4	0.28	.7795
Blame	41	30	−17.0	6.07	.0000	−14.3	3.04	.0024	−11.1	2.36	.0183	−10.0	1.88	.0601	−13.5	2.55	.0189
Control	23	43	−8.7	2.12	.0340	−0.1	0.02	.9840	−2.1	0.36	.7189	−1.4	−.26	.7949	0.6	0.12	.9045

In Table XVIII, which is a composite of several of the preceding tables, it can be seen that in nine cases (groups) out of eleven the significant difference in gains is accompanied by a significant difference in the S_I

TABLE XIV

SIGNIFICANCE OF DIFFERENCE IN GAINS, UNCORRECTED FOR INFLUENCE OF CA, MA, AND IS. TESTER A VS. TESTER B. GRADES VII, VIII, AND H.S.

Group	Tester A vs. Tester B			N	
	Diff.	k	P	A	B
VII					
P vs. P	31.1	4.82	.0000	26	29
B vs. B	12.1	2.05	.0404	29	30
C vs. C	7.3	1.43	.1527	30	22
VIII					
P vs. P	24.6	4.95	.0000	31	31
B vs. B	11.6	1.84	.0658	30	28
C vs. C	19.9	3.73	.0002	33	33
H.S.					
P vs. P	9.7	1.38	.1676	14	41
B vs. B	3.5	0.70	.4839	41	30
C vs. C	9.3	1.84	.0658	23	43

score; but in only two cases out of the eleven is this difference accompanied by a significant difference in the S_c score. Schematized, this appears thus:

SIGNIFICANT DIFFERENCE IN GAINS ACCOMPANIED BY:

	Sig. Diff. in S_e	Non-Sig. Diff. in S_e	Totals
Sig. Diff. in S_I	2	7	9
Non-Sig. Diff. in S_I	0	2	2
Totals	2	9	11

In Table XVIII the significant differences all have been italicized. The asterisk indicates that when the differences have been corrected for the influence of MA, CA, and Initial Score, they are significant (*cf.* Tables XV and XVII).

TABLE XV

SIGNIFICANCE OF DIFFERENCE IN GAINS, CORRECTED FOR INFLUENCE OF CA, MA, AND IS. TESTER A VS. TESTER B. GRADES VII, VIII, AND H.S.

	Tester A vs. Tester B			N	
Group	Diff.	k	P	A	B
VII					
P vs. P	35.2	6.07	*.0000*	26	29
B vs. B	17.5	3.13	*.0017*	29	30
C vs. C	13.5	2.23	.0257	30	22
VIII					
P vs. P	28.6	5.15	.0000	31	31
B vs. B	11.9	2.14	*.0324*	30	28
C vs. C	23.0	4.31	*.0000*	33	33
H.S.					
P vs. P	14.3	2.26	*.0238*	14	41
B vs. B	4.7	0.96	.3271	41	30
C vs. C	10.3	1.97	.0483	23	43

Recessional

In only 31 individual cases did subjects fail to make a better score on succeeding tests up to S_e. A few of these, evidently from examination of their papers, were ones who had misunderstood directions and had begun incorrectly, but later discovered their mistake and had begun over at S_1 or S_2. However, at S_e there was noted in many instances in individual cases an actual decrease in score from that of S_3. In all groups between S_1 and S_3 only 31 subjects, or 5.4 per cent, made a poorer score on succeeding tests than on the one immediately prior. On S_e for all groups, however, a total of 98 subjects or 17+ per cent, did more poorly than on S_3.

TABLE XVI

Significance of Difference in Gains, Uncorrected for Influence of CA, MA, and IS. Tester A vs. Tester B. Boys and Girls. Grades VII–VIII Combined

Group	N		Diff.	k	P
	A	B			
Boys					
P vs. P	29	31	28.3	5.55	.0000
B vs. B	29	26	9.4	1.62	.1052
C vs. C	28	26	7.3	1.38	.1676
Girls					
P vs. P	28	29	26.0	4.33	.0000
B vs. B	30	32	15.4	2.37	.0178
C vs. C	35	29	20.4	3.92	.0001

TABLE XVII

Significance of Difference in Gains, Corrected for Influence of CA, MA, and IS. Tester A vs. Tester B. Boys and Girls. Grades VII–VIII Combined

Group	N		Diff.	k	P
	A	B			
Boys					
P vs. P	29	31	32.5	6.33	.0000
B vs. B	29	26	11.3	2.12	.0340
C vs. C	28	26	9.3	1.74	.0819
Girls					
P vs. P	28	29	31.6	5.44	.0000
B vs. B	30	32	18.0	3.37	.0008
C vs. C	35	29	33.6	8.29	.0000

TABLE XVIII

COMPOSITE TABLE SHOWING GAINS AND DIFFERENCES IN GAINS (RAW). TESTER A VS. TESTER B

	Praise				Blame				Control			
	N	S_C	S_I	Gain	N	S_C	S_I	Gain	N	S_C	S_I	Gain
VII, Boys and girls												
Tester A	26	97.3	18.9	78.4	29	81.3	20.4	60.9	30	79.9	21.6	58.3
Tester B	29	78.1	30.6	47.5	30	83.1	34.3	48.8	22	83.1	32.1	51.0
Raw Diff.		*19.2*	*11.9*	31.1*		−1.8	*−13.9*	12.1*		−3.2	*−10.5*	7.3
VIII, Boys and Girls												
Tester A	31	98.3	25.6	72.7	30	97.7	26.6	71.1	33	97.2	33.3	63.9
Tester B	31	88.6	40.5	48.1	28	88.8	29.3	59.5	33	88.4	44.4	44.0
Raw Diff.		9.7	*−14.9*	24.6*		9.9	−2.7	11.6		8.8	*−11.1*	19.9*
VII–VIII, Boys												
Tester A	29	98.5	21.2	77.3	29	80.2	19.3	60.9	28	83.3	27.6	55.7
Tester B	31	81.9	32.9	49.0	26	81.2	29.7	51.5	26	86.3	37.9	48.4
Raw Diff.		16.6	11.7	29.3		−1.0	−10.4	9.4		−3.0	−10.3	7.3
VII–VIII, Girls												
Tester A	28	97.1	23.5	73.0	28	99.2	27.6	71.6	30	93.4	27.6	65.8
Tester B	29	86.4	38.9	47.6	32	89.8	33.6	56.2	29	86.2	40.8	45.4
Raw Diff.		10.7	*−15.3*	26.0*		9.4	−6.0	15.4*		7.2	*−13.2*	20.4*
H.S., Boys												
Tester A	14	104.0	32.9	71.1	41	107.4	32.9	74.5	23	98.9	33.0	65.9
Tester E	41	106.0	44.6	61.4	30	120.9	49.9	71.0	43	98.3	41.7	56.6
Raw Diff.		−2.0	*−11.7**	9.7*		*−13.5*	*−17.0*	3.5		0.6	−8.7	9.3*

X—Significant differences are *italicized*. See Tables VII, X, XI, XIV, XVI.
*—The asterisk indicates that when this difference has been corrected for the influence of CA, MA, and IS it is significant. See Tables XV, XVII.

Subjectively, it was observed that the announcement of the reward of one dollar brought not a few "Ah's" and "Oh's", and other comments, which were followed in many instances by quite feverish activity.

In order to determine whether or not the percentages of those receding on the last test was significantly larger than those receding on all tests up to the last, the probable error of difference between the two percentages was determined. The type formula of $PE_{diff} = \sqrt{(PE_1)^2 + (PE_2)^2}$ was employed, the PE in each case being a derivation of the proportion probable error, $PE_p = .6745 \sqrt{\frac{pq}{n}}$.

The difference in percentage was 11.67, and the probable error of this difference was 1.23, showing that there is a significantly larger proportion receding on the last test than on all the other tests combined up to the last test.

The Correction

Returning to our earlier discussion on the question of equating for various variables, some interesting phenomena might be noted. What may have been lost had actual physical equating of the groups been effected, is, of course, almost wholly conjectural. On the other hand, the fact cannot be evaded that any estimate of a parameter, whatever its refinements may be, still remains only an estimate. In this experiment, all the cases have been retained and a statistical procedure has been employed to correct for the influences of MA, CA, and Initial Score. It should follow that inquiry be made as to the effects of the corrections or, perhaps more pertinently, the effects of the variables of MA, CA, and Initial Score.

Tables VIII, IX (pp. 38–39), XIV, XV, XVI, and XVII (pp. 43–45) show the gains uncorrected and the gains corrected for the influence of MA, CA, and Initial Score. It will be noted that in some instances the disparity between the two gains seems of appreciable magnitude. However, further analysis shows that P values for uncorrected gains have been altered significantly with correcting in only a few instances. These several cases are to be found in cases of Tester A vs. Tester B. From Table XIV it will be seen that for grade VII in the case of the Controls, P was equal to .1527, while in Table XV, where the gains had been corrected, P becomes .0257. From the same tables, there is similarly, for grade VIII, a Blame change for P from .0658 to .0324; and for high school, Praise vs. Praise, a change from .1676 to .0238.

Where the sexes and the different testers are compared, Tables XVI and XVII, for boys, Blame, P changes from .1052 to .0340.

What seems apparent from analysis of the data in this study is that there was little or negligible effect of the variables MA, CA, and Initial Score; and that any equating process could have been dispensed with without in any way vitiating the results. Indeed, this is foreshadowed by the correlations obtained at the outset (Tables III and IV, pp. 31 and 32).

Physiological

Figure 15 shows graphically the $\bar{x}$ scores for pulse rate of the three college groups on the separate tests. There were no significant differences from test to test for any of the groups; but the curves, however, show several traits in common.

There is a rise from the preliminary period, S_p, to S_I, a slight decrease in rate at S_1, then a gradual rise with a more pronounced incidence from S_3 to S_e, and a final falling off at S_F, which approaches the point reached at S_p.

For the breathing ratio, Figure 16 portrays graphically the $\bar{x}$ scores of the groups on the separate tests. This was the "I-factor" as described by Woodworth (83).[1] As was the case with the pulse rate, here again there appear no significant differences between groups, but the trend of the curves is interesting. Contrary to the pulse rate, there is a tendency for the inspiration-expiration ratio to lessen gradually to S_3; but similarly, as was the case with the pulse rate, after the announcement of the reward at S_3 (that is, just before the beginning of the final test), there is a distinct rise to S_e, followed by subsidence to S_F.

The data on blood pressure did not lend themselves to definite or quantitative handling and shall not be presented here. However, certain trends did seem to be manifest, as where there was a drop at "Stop," and an increase again at "Begin." This was a fairly general condition to prevail, but not always regular. Often the indication of a change was slight or similar to the following: [2].

[1] For greater ease in calculating, this procedure was modified a trifle here by multiplying the factor by 100.

[2] The method employed of selecting an arbitrary pressure, as was brought out earlier, gives neither a systolic nor a diastolic pressure, but will indicate a *change* in the circulatory system. This was originally brought out by J. A. Larson, *Lying and Its Detection*, 1932; Christian A. Ruckmick, *The Psychology of Feeling and Emotion* (New York: McGraw Hill Co., 1936), pp. 291 and 309, speaks of it as an applicable technique. Others, as Harold V. Gaskill, "The Objective Measurement of Emotional Reactions," *Genet. Psychol. Monogr.*, Vol. XIV, No. 3, 1935, p. 204, have employed this principle using pressure reducers.

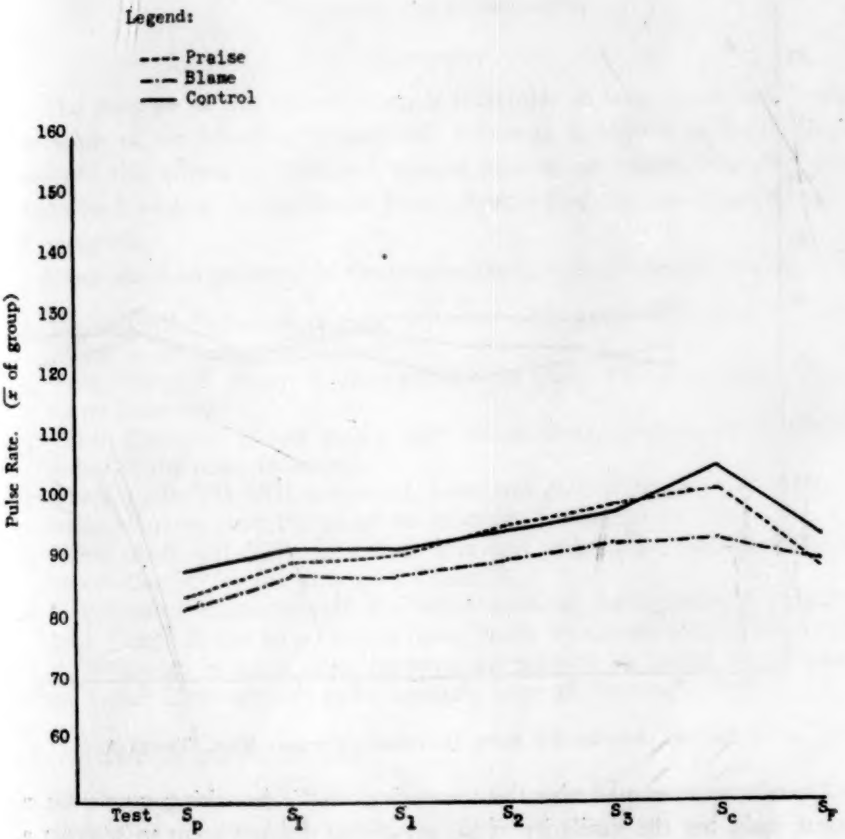

FIG. 15. AVERAGE PULSE RATE PER MINUTE. COLLEGE MEN. TESTER A

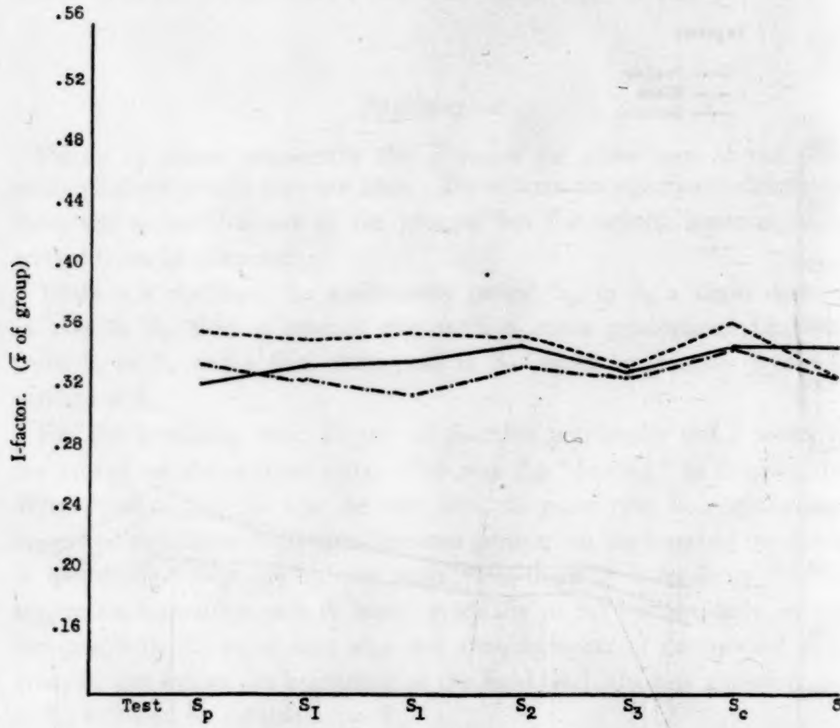

FIG. 16. AVERAGE I/E RATIO (I-FACTOR). COLLEGE MEN. TESTER A

There was attempted also the measuring of the breathing curves on a linear basis, but the similarity of the samplings did not seem to warrant a continuation of this procedure.

CHAPTER IV

Summary and Conclusions

Summary

The purpose of this investigation is threefold: to seek an answer to the question of the effects of Praise and Blame as incentives in learning; to observe the effects of different testers; and to determine the attendant emotional aspects, as evidenced from physiological changes recorded on a kymograph.

From the data gathered in the present study, it is practically certain that:

1. Praise with Tester A is more effective with grade VII than is either Blame or no incentive.
2. With Tester B, Blame is more effective in grade VIII than either Praise or no incentive.
3. With Tester B, Blame at the high school level is more effective than either of the other incentives.
4. With grade VII–VIII combined, boys and girls differentiated, Praise is more effective than Blame or no incentive for boys with Tester A.
5. With grade VII–VIII combined, boys and girls differentiated, Blame is more effective for the girls with Tester B.
6. A difference exists between the initial scores of the subjects of Tester A and Tester B, the larger scores being made by the subjects of Tester B.
7. A difference in gains exists between the subjects of Tester A and those of Tester B, the greater gains being in favor of Tester A.

It can also be concluded that:

1. With sexes compared, incentive group vs. incentive group, the differences between boys and girls are not of statistical significance.
2. There is a continued increase in score of the various groups from test to test up to the final test.
3. There is a statistically significant suggestion that a loss on the final test from that of the test immediately preceding is not due to chance factors.
4. There appears a tendency for Praise to be more effective with Tester A, and Blame for Tester B.
5. There appear no physiological patterns to distinguish one form of incentive from the other.
6. The curves for pulse rate for Praise, Blame, and Control groups follow the same pattern. There is the tendency for the pulse rate to increase slightly from test to test with a marked rise from S_3 to S_c, followed by a drop to S_F.

7. The pattern of the curves for the I/E ratio (I-factor) is closely the same for the three groups. There is a tendency for the ratio to decrease from test to test, with a distinct rise from S_3 to S_e, followed by a drop to S_F.

Conclusions

This investigation was not concerned with testing the effects of Praise or Blame as psychological opposites, but rather their effectiveness, *per se,* in producing a change in the learning situation. The statements of Praise or Reproof were administered rather much as would be done in any routine classroom situation of like nature; but with the precaution that no external factors of either tester would bias the results. The same clothes were worn on each testing occasion and the directions and comments were rehearsed so as to be delivered with identical intonation and inflection and without extraneous clues (*vide* p. 27).

From the data as they are found, the conclusion is arrived at that neither Praise nor Blame can be singled out as being more effective, the one over the other. On the one hand Tester A found, in grade VII–VIII combined, Praise to be more effective for boys, but found no differences for girls; on the other hand, in grade VII–VIII combined, girls (but not boys), and in high school, boys, Tester B found Blame to be the more effective. Tester A found no differences at the high school or college levels; and in no case did Tester A find Blame more effective. Similarly, in no case did Tester B find Praise more effective.

It is more than just apparent that the test situation, and specifically *the tester,* plays a deciding rôle in the results obtained. In every group but two the initial scores made by Tester B were significantly higher than those made by Tester A. In the larger majority of instances the gains made by Tester A were significantly larger than those for Tester B.

It must be concluded also that the nature of the stimulus is most important; for if it be non-effective on the one hand, it may be so intense as to inhibit improvement on the other. Which is to say, if there appears no evidence to support either Praise or Blame as more effective incentives to learning, there does seem to be evidence to point to the effect of overstimulation lessening improvement. Up to the final test, all tests included, only five per cent of the subjects made a poorer score on a succeeding test, but on the final (reward) test alone, 17 per cent lost ground.

Implications for Education

It would seem from these data that in standardizing test procedure that the *who* is testing should be of some concern. The data of this study have shown that without exception the subjects made a higher initial score with one tester than with the other. The differences were statistically significant in all but two groups. These differences developed in the face of strictly standardized test procedure, rigid adherence thereto, and control of the objective features of each tester.

If standardized test scores are to be meaningful, it would appear that the following of standard test procedure alone is not sufficient to insure comparable results. What is indicated is a calibration of test scores in terms of the tester, or a calibration of testers.

Differences as large as were found here might have serious consequences to individual pupils and to a school system. If a system employs an X, Y, and Z division of pupils, for instance, it is quite conceivable then, that pupils will be misplaced; too high, perhaps, in some instances, and too low in others. The problem of the proper placement of pupils and the harmful effects of misplacement have received too much attention elsewhere to warrant here more than an indication of a possible cause of maladjustment in school.

Equally vital appears the question of the " best " incentive to be employed in order to bring out the most in a pupil, in order to get him to *perform* maximally. From the data of this study, one can find no evidence favoring either Praise or Blame as fulfilling this rôle. The evidence here submitted, certainly, does not substantiate the popular belief of the necessity and efficacy of Praise as a prime motivator. To employ Praise in the belief or in the confidence that a pupil will be made thus to react, or to produce, maximally, is to proceed upon false assumptions, is to build on a foundation of clay.

BIBLIOGRAPHY

1. ABEL, LORRAINE B. The effect of shift in motivation upon learning of a sensori-motor task. *Arch. Psychol., N. Y.*, 1936, **205**, 57 pp.
2. ADAMS, J. Praise as an educational instrument. *J. Educ., London*, 1934, **66**, 506–507.
3. ANDERSON, H. H. Motivation of young children: further studies in success and failure, praise and blame. *Child Developm.*, 1936, **7**, 125–143.
4. ——— and SMITH, RUTH SLOAN. Motivation of young children: The constancy of certain behavior patterns. *J. exper. Educ.*, 1933–34, **2**, 138–160.
5. ARPS, G. F. Attitude as a determinant in spelling efficiency in immediate and delayed recall. *J. educ. Psychol.*, 1915, **6**, 409–418.
6. ——— A preliminary report of work with knowledge of results vs. work without knowledge of results. *Psychol. Rev.*, 1917, **24**, 449–455.
7. ——— Work with knowledge of results vs. work without knowledge of results. *Psychol. Monogr.*, 1920, **28**, No. 3.

8. BENTON, A. L. Influence of incentives upon intelligence test scores of school children. *J. genet. Psychol.*, 1936, **49**, 494–497.
9. BILLS, A. G. General experimental psychology. New York: Longmans Green Co., 1934, 620 pp., pp. 258–274.
10. BINET, A., and VASCHIDE, E. Expérience de forces musculaire et du fond chez les jeunes garçons. *Année Psychol.*, 1897, **4**, 15–63.
11. BIRD, CHARLES. Effective study habits. New York: Century Co., 1931, p. 17.
12. ——— Personality factors in learning. *Person. J.*, 1927, **6**, 56–59.
13. BLANKENSHIP, M. V., and HUMES, J. F. The effect of praise and reproof upon memory span performance. *Amer. J. Psychol.*, 1938, **51**, 527–531.
14. BLISSENBACH, M. V. Praise as a factor in religious education. *Cath. Sch. J.*, 1938, 38, 199–202.
15. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929–1931, Vol. IV, 113–134.
16. BOOK, WILLIAM F. Economy and technique of learning. Boston: D. C. Heath & Co., 1927, pp. 147–150.
17. ——— and NORVELL, L. The will to learn. *J. genet. Psychol.*, 1922, **29**, 305–362.
18. BRENNER, B. Effect of immediate and delayed praise and blame upon learning and recall. *Teach. Coll. Contr. Educ.*, 1934, No. 620.
19. BRIGGS, THOMAS H. Praise and censure as incentives. *Sch. & Soc.*, 1927, **26**, 596–598.
20. ——— Sarcasm. *Sch. Rev.*, 1928, **36**, 685–695.
21. CHAPMAN, I., and FEDER, R. B. The effect of external incentives on improvement. *J. educ. Psychol.*, 1917, **8**, 469–475.
22. CHASE, LUCILLE. Motivation of young children: An experimental study of the influence of certain types of external incentives upon the performance of a task. *Univ. Iowa Stud. Child Welf.*, 1932, Vol. **5**, No. 3.
23. COHEN, A. A study of incentives under schoolroom conditions. M.A. Thesis, Columbia Univ., 1927.
24. COLE, LOUELLA. Teaching in the elementary school. New York: Farrar & Rinehart, 1939, xvi, 518 pp., pp. 179–180.
25. DAVIS, R. A., and BALLARD, C. R. Effectiveness of various types of classroom incentives. *Educ. Meth.*, 1932, **12**, 134–145.
26. DEPUTY, E. B. Knowledge of success as a motivating factor in college work. *J. educ. Res.*, 1929, **20**, 327–334.
27. DISERENS, CHARLES M., and VAUGHN, JAMES. The experimental psychology of motivation. *Psychol. Bull.*, 1931, **28**, 15–65.
28. EATON, M. T. Effect of praise, reproof and exercise upon muscular exercise. *J. exp. Educ.*, 1933, **2**, 44–59.
29. ELWELL, J. L., and GRINDLEY, G. C. The effects of knowledge of results on learning and performance. *Brit. J. Psychol.*, 1938, **29**, 39–54.
30. FINAN, JOHN L. (assisted by Taylor, Lloyd F.). Quantitative studies in motivation: I. Strength of conditioning in rats under varying degrees of hunger. *J. comp. Psychol.*, 1940, Vol. 29, No. 1, 119–134.
31. FORLANO, H., and AXELROD, C. The effect of repeated praise or blame on the performance of introverts and extroverts. *J. educ. Psychol.*, 1937, **28**, 92–100.
32. GARRISON, K. C. Motivation and learning. *Peabody Reflector*, 1934, **17**, 680–690. *These data appear also in* Garrison, S. C., and Garrison, K. C. *Fundamentals of Psychology in Secondary Education*. New York: Prentice-Hall, 1936, pp. 148–170.
33. GASKILL, HAROLD V. The objective measurement of emotional reactions. *Genet. Psychol. Monogr.*, 1935, Vol. XIV, No. 3, p. 204.
34. GATES, ARTHUR I. The improvement of reading. New York: The Macmillan Co., 1935, p. 571.
35. GATES, G. S., and RISSLAND, L. The effect of encouragement and discouragement upon performance. *J. educ. Psychol.*, 1923, **14**, 21–22.
36. GILCHRIST, E. P. The extent to which praise and reproof affect a pupil's work. *Sch. & Soc.*, 1916, **4**, 872–874.
37. GOULDEN, C. H. Methods of statistical analysis. New York: John Wiley & Sons., 1930, pp. 10–12.
38. GRIFFITH, COLEMAN R. Psychology of coaching. New York: Scribner's Sons, 1929, p. 198.
39. HARRIS, CHARLOTTE. Right: an experiment with after-effects. *J. exp. Educ.*, 1935, 4, 67–72.
40. HOLZINGER, KARL J. Statistical methods for students in education. New York: Ginn & Co., 1928, chap. XV, "Partial and multiple correlation," pp. 283–316; *also*, pp. 243, 248.

41. HURLOCK, ELIZABETH B. The value of praise and reproof as incentives for children. *Arch. Psychol., N. Y.*, 1924, No. 71.
42. ——— An evaluation of certain incentives used in school work. *J. educ. Psychol.*, 1925, **16**, 145–159.
43. ——— The effects of incentives upon the constancy of the I.Q. *J. genet. Psychol.*, 1925, **32**, 422–434.
44. ——— The use of group rivalry as an incentive. *J. abnorm. (soc.) Psychol.*, 1927, **22**, 278–290.
45. ——— The psychology of incentives. *J. soc. Psychol.*, 1931, **2**, 261–290.
46. IRWIN, M. E. Motivation, emotional responses, maturation, intelligence and individual differences. *Rev. educ. Res.*, 1936, **6**, 300–309.
47. JUDD, C. H. Practice without knowledge of results. *Psychol. Rev. Monogr. Suppl.*, 1906, **7**, 185–199.
48. KIRBY, T. J. Practice in the case of school children. Columbia Univ., *Teach. Coll. Contr. Educ.*, 1913, series No. 58.
49. KNIGHT, F. B., and REMMERS, H. G. Fluctuations in mental production when motivation is the chief variable. *J. appl. Psychol.*, 1923, **7**, 209–223.
50. LAIRD, D. A. Changes in motor control and individual variations under influence of "razzing." *J. exp. Psychol.*, 1923, **6**, 236–246.
51. ——— How the high school student responds to different incentives to work. *J. genet. Psychol.*, 1923, **30**, 358–365.
52. ——— How the college student responds to different incentives to work. *J. genet. Psychol.*, 1923, **30**, 366–370.
53. LARSON, J. A. Lying and its detection. Chicago: Univ. Chic. Press, 1932.
54. LASHLEY, K. S. Nervous mechanisms in learning. Chap. 10 *in* Murchison, Carl, *Handbook of General Experimental Psychology.* Worcester: Clark Univ. Press, 1934, pp. 456–496.
55. LEEPER, ROBERT. The rôle of motivation in learning: A study of the phenomenon of different motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
56. ——— A study of a neglected portion of the field of learning in the development of sensory organization. *J. genet. Psychol.*, 1935, **46**, 41–75.
57. LEUBA, C. J. A preliminary experiment to quantify an incentive and its effect. *J. abnorm. (soc.) Psychol.*, 1930, **25**, 275–288.
58. MACLEARIE, E. C. Approval program. *Jr.-Sr. H. Sch. Clearing House*, 1934, **9**, 106–107.
59. MALLER, J. B. Co-operation and competition. *Teach. Coll. Contr. Educ.*, Columbia Univ., 1927, No. 384.
60. ——— and ZUBIN, J. The effect of motivation upon intelligence test scores. *J. genet. Psychol.*, 1932, **41**, 136–151.
61. MURPHY, G., MURPHY, L. B., and NEWCOMB, T. M. Experimental social psychology (revised ed.). New York: Harper & Bros., 1937, Chap. VII, *Characteristic social behavior of children in our culture: aggression and competition*, pp. 376–508.
62. NELSON, J. G. A preliminary report on some uses of the psychological test situation for studying personality differences, *in* Thomas, D. S., and associates: *Some New Techniques for Studying Social Behavior. Child Developm. Monogr.*, Columbia Univ., 1922, No. 1, pp. 1–23.
63. Otis Quick-Scoring Tests of Mental Ability, Gamma, Form Am. Yonkers: World Book Co.
64. REAM, J. A tip on managing people. *J. appl. Psychol.*, 1924, **8**, 357–361.
65. RUCKMICK, CHRISTIAN A. The psychology of feeling and emotion. New York: McGraw Hill & Co., 1936, pp. 291, 309.
66. RYANS, DAVID G. An experimental attempt to analyze persistent behavior: I. Measuring traits presumed to involve persistence. *J. genet. Psychol.*, 1938, **19**, 333–353.
67. SCOTT, W. D. Increasing human efficiency in business. New York: The Macmillan Co., 1923, p. 10.
68. SEARS, ROBERT R. Initiation of the repression sequence by experienced failure. *J. exp. Psychol.*, 1937, **20**, 570–580.
69. SIMS, V. M. The relative influence of two types of motivation on improvement. *J. educ. Psychol.*, 1928, **19**, 480–484.
70. SPENCER, L. T. The effect of practice without knowledge of results. *Amer. J. Psychol.*, 1923, **34**, 107–111.
71. SULLIVAN, E. B. Attitude in relation to learning. *Psychol. Monogr.*, 1927, **36**, 149 pp.
72. SYMONDS, P. M., and CHASE, D. H. Practice vs. motivation. *J. educ. Psychol.*, 1929, **20**, 19–35.

73. TERMAN, L. M., and MERRILL, MAUD A. Measuring intelligence. New York: Houghton Mifflin Co., 1937, p. 31.
74. TOLMAN, E. C. Purposive behavior in animals and man. New York: Century Co., 1932, xiv, 463 pp.
75. ——— HALL, C. S., and BRETNALL, E. P. A disproof of the law of effect and a substitution of the laws of emphasis, motivation, and disruption. *J. exp. Psychol.*, 1932, **15**, 601–614.
76. TRELOAR, ALAN E. Elements of statistical reasoning. New York: John Wiley & Sons, 1939, p. 77, 150; Appendix 1, p. 239.
77. TRIPLETT, N. The dynamogenic factors in pace-making and competition. *Amer. J. Psychol.*, 1897–1898, **9**, 507–533.
78. TROW, W. C. Motivation, emotional responses, maturation, intelligence, and individual differences. *Rev. educ. Res.*, 1939, **9**, 285–294.
79. WARDEN, C. J., and COHEN, A. A study of certain incentives applied under schoolroom conditions. *J. genet. Psychol.*, 1931, **39**, 320–327.
80. WATTS, F. An introduction to the psychological problems of industry. New York: The Macmillan Co., 1924, 240 pp.
81. WOLF, THETA H. The effect of praise and competition on the persistent behavior of kindergarten children. *Instit. Child Welf. Monogr.*, Series No. 15, Univ. of Minn. Press, 1938, 138 pp.
82. WOOD, T. W. The effect of approbation and reproof on the mastery of nonsense syllables. *J. appl. Psychol.*, 1934, **18**, 657–664.
83. WOODWORTH, ROBERT S. Experimental psychology. New York: Henry Holt & Co., 1938, pp. 262–266.
84. YOUNG, P. T. Motivation of behavior. New York: John Wiley & Co., 1936, 562 pp., pp. 388–416.